A functional analysis of a family of proteins implicated in *Trypanosoma brucei* lifecycle progression

Samuel Dean

PhD

The University of Edinburgh

2008

Abstract

Bloodstream trypanosomes initiate differentiation to procyclic forms in response to a citrate/ *cis*-aconitate (CCA) signal. A cell line was previously selected ("defective in differentiation-clone 1"; DiD1) that was unable to differentiate to procyclic forms (Tasker *et al.* (2000)). Additionally, expression profiling of this line in comparison to the parental line by macroarray hybridisation identified two differentially-expressed transcripts from an 8 gene cluster of highly homologous genes we named PAD genes (Proteins Associated with Differentiation). Members of this family show distinct expression profiles throughout the trypanosome lifecycle at both the mRNA and protein level, and are localised to the cell surface membrane of the cell. At least 1 member of the family (PAD1) shows stumpy form specific RNA and protein expression, representing the first useful molecular marker for this stage, and exhibits biochemical specificity for citrate. Additionally, another member of this family (PAD2) is upregulated in response to low temperature, a condition reported to cause hypersensitivity to CCA. Finally, RNAi mediated ablation of the PAD gene transcripts compromised the capacity of stumpy form trypanosomes to differentiate to the procyclic form in response to CCA. These combined expression, cytological, reverse-genetic and biochemical data make PAD proteins excellent candidates for recognition of the signal to initiate differentiation in response to CCA.

Samuel Dean

Matric: s9812834

PhD student

Table of Contents

Title page i

Abstract ii

Table of contents iii

List of Figures xii

List of Tables xv

Declaration xvi

Achnowledgments xvii

Abbreviations xviii

1 Introduction 1

1.1 Evolution of *Trypanosoma brucei spp.* 3

1.2 African Sleeping Sickness 3

1.2.1 Treatments 4

1.3 Lifecycle 9

1.3.1 Bloodstream forms 9

1.3.2 Insect forms 12

1.3.3 Stumpy to Procyclic form differentiation 13

1.3.4 Mesocyclic form 15

1.3.5 Epimastigote form 15

1.3.6 Metacyclic form 17

1.4 Chemicals reported to induce slender to stumpy form differentiation 20

1.4.1 The role of cAMP signalling 20

1.4.2 Thiazolidinediones 22

1.4.3 DL-α-difluoromethylornithine (DFMO) 23

1.5 Conditions reported to induce stumpy to procyclic form differentiation 24

1.5.1 Incubation at 27°C 24

1.5.2 Incubation at 27°C and the addition of tri-carboxylates 24

1.5.3 Mild acid treatment 26

1.5.4 Proteolytic stress 27

1.5.5 Replacement of glucose with glycerol in the media 27

1.6 The relationship between cell cycle position, life cycle position, and competence to differentiate 28

1.7 Transporters 30

1.7.1 Adenosine binding cassette (ABC) transporters 31

1.7.2 Equilibrative nucleoside transporters (ENT) 32

1.7.3 Major Facilitator Superfamily (MFS) 32

1.7.4 Channels 34

1.8 DiD1 generation and identification of the Proteins Associated with Differentiation (PAD) gene family 36

1.9 Aims and objectives 40

2 Materials and Methods 41

2.1 Trypanosomes 42

2.1.1 Strains and cell lines 42

2.1.2 Culturing 43

2.1.3 Purification of trypanosome cells from blood 44

2.1.4 Differentiation 44

2.1.5 Transfection 45

2.1.6 Inducing cold shock 47

2.2 Bacteria 48

2.2.1 Strains 48

2.2.2 Transformation 48

2.3 Molecular Biology techniques 48

2.3.1 Polymerase Chain Reaction 48

2.3.2 Preparation and purification of DNA 49

2.3.3 DNA restriction enzyme digestion 50

2.3.4 DNA ligation 50

2.3.5 DNA sequencing 50

2.4 Northern analysis 51

2.4.1 Riboprobe preparation 51

2.4.2 Resolving trypanosome RNA on a denaturing agarose gel 51
2.4.3 Blotting trypanosome RNA 52
2.4.4 Probe hybridization and detection 52
2.5 Southern Analysis 53
2.5.1 DNA probe synthesis 53
2.5.2 Resolving trypanosome genomic DNA on an agarose gel 53
2.5.3 Genomic DNA blotting 53
2.5.4 Probe hybridization and detection 53
2.6 Western analysis 54
2.6.1 Preparation of protein sample 54
2.6.2 Preparation of SDS gel 54
2.6.3 Resolving proteins on an SDS polyacrylamide gel 54
2.6.4 Blotting proteins 55
2.6.5 Quantitative Western blots 56
2.6.6 Western blot of PAD proteins 57
2.7 Immunofluorescence 57
2.7.1 Methanol fixation 58
2.7.2 Paraformaldehyde fixation 58
2.8 Flow cytometry 59
2.8.1 Fixing cells 59
2.8.2 Blocking and staining 59
2.8.3 Use of the FACS machine 59
2.8.4 Analysis of data 60
2.9 Citrate transport assay 60
2.9.1 Synthesis of PAD1 RNA 60
2.9.2 Preparation of *Xenopus* oocytes for micro-injection 60
2.9.3 Microinjection of *Xenopus laevis* oocytes 61
2.9.4 Citrate transport assay 61
2.9.5 Analysis of data 61
2.10 Plasmid constructs used in this study 61

2.10.1 pHD451 62

2.10.2 pHD617 and pHD617 (PuroR) 62

2.10.3 pHD449 62

2.10.4 pALC14 63

2.10.5 Gene deletion constructs 63

2.10.6 pGEM T easy 64

2.10.7 pGem-He-Juel 64

3 A Bioinformatics analysis of the PAD array 65

3.1 The genomic context of the PAD genes 67

3.2 Relationships within the PAD family 76

3.3 An analysis of the PAD array intergenic regions 82

3.3.1 Predicting the positions of the PAD *trans*-splicing sites 82

3.3.2 Aligning the PAD UTRs 83

3.3.3 An analysis of the repetitive elements within the PAD intergenic regions 83

3.3.4 Predicting the secondary structure of the PAD 3'UTRs 91

3.4 Detecting conserved domains within PAD proteins 94

3.4.1 Pfam analysis 94

3.4.2 CDD analysis 95

3.4.3 Prosite analysis 95

3.5 Predicting the localisation of the PAD proteins 97

3.6 Predicting the secondary structure of the PAD proteins 100

3.7 Predicting the three-dimensional structure of the PAD proteins 103

3.8 Using PSI-BLAST to help elucidate the PAD proteins' function 103

4 Characterisation of the DiD1 cell line 106

4.1 A Northern analysis to compare PAD genes in the DiD1 and parental cell lines 107

4.2 A Southern analysis to detect PAD array genome rearrangements 110

4.3 A sequencing analysis to compare the PAD1 3' intergenic region in the DiD1 and parental cell lines 113

4.4 A sequencing analysis to determine the presence of mutations of PAD genes in the DiD1 cell line 116

4.5 Ectopic expression of PAD1 to recapitulate the DiD1 differentiation defect 120

4.5.1 Over expression of PAD1 120

4.6 A Western to determine whether PAD proteins are upregulated in the DiD1 cell line 123

4.7 A Northern analysis to determine whether there is a global upregulation of stumpy-enriched transcripts in the DiD1 cell line 126

4.7.1 TbPTP1 126

4.7.2 ESAG-K9 and glutamine synthase 127

5 A molecular characterisation of the PAD array 129

5.1 A Northern analysis to determine the RNA expression profile of the PAD genes 130

5.2 Localisation of the PAD1 protein 136

5.2.1 Localisation of an ectopically expressed, epitope tagged copy of PAD1 136

5.2.2 Localisation of the endogenous PAD1 protein 143

5.3 Optimising detection of PAD1 by Western analysis using an ectopically expressed, epitope tagged copy of PAD1 148

5.4 PAD1 is a stumpy form specific protein 151

5.4.1 Western analysis 152

5.4.2 Immunofluorescence 152

5.4.3 Flow cytometry analysis 155

5.5 Investigating PAD2 expression 157

5.5.1 Designing antigenic peptides to generate α-PAD2 antibodies 157

5.5.2 A Western blot analysis to determine PAD2 expression through the trypanosome lifecycle 160

5.6 Analysing the lifecycle expression of the entire PAD family 162

5.6.1 Generating antigenic peptides to generate antibodies recognising all PAD array members 162

5.6.2 Expression of unidentified PAD proteins through the trypanosome lifecycle 165

5.6.3 Identification of the bands detected by the α-array antibodies 168

5.7 Expression of PAD1 during synchronous differentiation to procyclic forms 170

5.8 Expression of PAD2 through differentiation 172

5.9 Expression of the entire PAD family of proteins through differentiation 174

5.10 Expression of PAD1 is associated with expression of EP procyclin during differentiation of pleomorphic cells 177

5.11 PAD expression in monomorphic slender cells 180

5.11.1 The affect of differentiation conditions on PAD1 protein expression in monomorphs 180

5.12 Investigating the kinetics of cold shock 183

5.13 Cold shock does not cause increased expression of PAD1 mRNA transcript 186

5.14 There was a non-reproducible up-regulation of PAD1 protein in response to cold-shock 188

5.14.1 The effect upon PAD1 expression of varying the degree of parasitaemia and the temperature of cold-shock 191

5.14.2 The effect upon PAD1 expression of cold shock in combination with foetal calf serum (FCS) deprivation 193

5.14.3 The effect of cold shock in combination with pH variation upon PAD1 expression 195

5.15 The effect of cold shock upon protein expression of other PAD members 197

5.15.1 The effect of cold shock upon the PAD family protein expression 197

5.15.2 The affect of cold shock in combination with foetal calf serum (FCS) deprivation upon PAD1 expression 199

5.15.3 The affect of pH and cold shock upon PAD array protein expression 201

5.16 The affect of cold shock upon PAD2 expression 203

5.17 Localisation of PAD2 in stumpy form cells 205

5.17.1 PAD2 localisation in stumpy forms at 37°C 205

5.17.1 PAD2 localisation in stumpy forms at 37°C 208

5.18 Localisation of PAD2 in procyclic form cells 210

5.19 Conclusion 212

6 An analysis of PAD family function 213

6.1 The effect of ectopic expression of PAD1 upon differentiation 214

6.1.1 The effect of expression of an ectopic copy of PAD1 upon the sensitivity to *cis*-aconitate at 20°C and 37°C 215

6.1.2 The affect of incubating cells with different concentrations of *cis*-aconitate during cold shock and ectopic expression of PAD1 219

6.2 Using RNA interference to ablate PAD1 in monomorphic slender cells 222

6.2.1 Generating the RNAi construct 222

6.2.2 The effect of inducing RNAi upon PAD1 225

6.3 Using RNA interference to ablate the PAD array 227

6.3.1 Generating the RNAi construct 227

6.3.2 The affect of inducing RNAi upon PAD protein and RNA 227

6.3.3 The effect of inducing RNAi upon capacity for differentiation 228

6.4 Generating transgenic knockout cell lines without the PAD1 gene 230

6.4.1 Making the constructs to remove PAD1 230

6.4.2 Analysing the transgenic cell lines 233

6.5 Generating transgenic PAD array null mutants 239

6.5.1 Generation of PAD array deletion constructs 239

6.5.2 Analysing the transgenic cell lines 240

6.6 RNAi mediated ablation of PAD family transcripts in pleomorphic cells 241

6.6.1 RNAi mediated ablated transfection of PAD family transcripts using transient transfection 241

6.6.2 Stable RNAi of the PAD array 251

6.6.3 The effect of PAD RNAi upon cold shock induced EP procyclin upregulation 257

6.7 Response of the RNAi cell line to an alternative differentiation trigger 259

6.8 Expression of PAD1 in Xenopus laevis oocytes to determine citrate transport activity 261

6.8.1 Optimisation of assay 262

6.8.2 PAD1 is able to transport citrate when expressed in Xenopus laevis 262

6.9 Conclusion 265

7 Discussion 266

7.1 Bioinformatics evidence 267

7.2 Analysis of the DiD1 defect 267

7.3 Expression of PAD1 in Xenopus laevis oocytes 269

7.4 Expression of PAD genes 269

7.4.1 PAD lifecycle expression and localisation 270

7.4.2 The affect of cold shock upon PAD2 271

7.5 Phenotypic analyses 273

7.5.1 PAD ablation 273

7.5.2 PAD ectopic expression 275

7.6 Future directions 276

7.6.1 Applications of a stumpy specific protein 276

7.6.2 Further dissection of the role of PAD proteins in promoting differentiation 277

7.6.3 The role of other PAD family members 278

7.6.4 Further analysis of the *cis*-aconitate mediated differentiation pathway 278

7.6.5 Citrate/ *cis*-aconitate uptake assays 279

7.7 Insights into trypanosome biology from this work 279

Appendix A **283**

Northern blot solutions 284

Southern blot solutions 285

Western blot solutions 286

Immunofluorescence solutions 287

Miscellaneous solutions 288

Small scale plasmid preperation solutions 289

Antibody concentrations 290

A list of primers used in this study 291

Plasmid maps 294

An alignment of PAD1 with the closest ortholog in *Trypanosoma cruzi* 295

A clustalW alignment of PAD1 with the closest ortholog in *Leishmania major* 296

A list of *T. cruzi* PAD1 orthologs 297

PAD array intergenic alignments 298-300

PAD protein secondary structures. 301

A PAD array protein alignment. 302

PAD deletion construct target sites and plasmid maps 303

An alignment of target sequence A 304

An alignment of target sequence B 305

An alignment of target sequence C 306

An alignment of target sequence D 307

An alignment of target sequence E 308

An alignment of target sequence F 309

Bibliography **310**

List of Figures

Figure 1.2 1 The structure of anti- HAT drugs 8

Figure 1.3 1 The trypanosome lifecycle 19

Figure 1.8 1 Selection of the DiD1 cell line 39

Figure 3.1 1 The genomic context of the PAD array 70

Figure 3.1 2 Synteny in the PAD region of the trypanosomatid genomes 71

Figure 3.1 3 A protein alignment of PAD1 and the closest orthologs in L. major and T. cruzi 72

Figure 3.1 4 The closest orthologs to the PAD genes in L. major are members of a 6 gene cluster 73

Figure 3.1 5 Close orthologs to the PAD genes in T. cruzi are members of a cluster containing at least 4 genes 74

Figure 3.1 6 The L. major and T. cruzi orthologous PAD arrays 75

Figure 3.2 1 The relationship between the PAD protein sequences 78

Figure 3.2 2 The relationship between the PAD gene nucleotide sequences 81

Figure 3.3 1 The positions and alignments of the PAD predicted UTRs 84

Figure 3.3 2 A comparison of the predicted PAD 3' UTRs. 87

Figure 3.3 3 Analysis of the repetitive elements within the PAD array intergenic regions 90

Figure 3.3 4 The predicted secondary structure of elements within the PAD 3' UTR as predicted by Sfold 93

Figure 3.4 1 A schematic showing the mechanisms of transport for major facilitator superfamily transporters 96

Figure 3.6 1 The secondary structure prediction of the PAD proteins 102

Figure 4.1 1 Expression of PAD genes in parental and DiD1 cells 109

Figure 4.2 1 A Southern blot to determine whether any genomic rearrangements have occurred during the selection of the DiD1 cell line 112

Figure 4.3 1 A sequencing analysis of the PAD1 3' UTR 115

Figure 4.4 1 An alignment of the parental, DiD1 and genome reference strain PAD1 encoded proteins 118

Figure 4.4 2 An alignment of the parental, DiD1 and genome reference strain PAD2 encoded proteins 119

Figure 4.5 1 The effect of PAD1 upon differentiation in response to cis-aconitate 122

Figure 4.6 1 PAD proteins are expressed in DiD1 cells 125

Figure 4.7 1 There is not a global stumpy form transcript-upregulation 128

Figure 5.1 1 The riboprobes designed to detect PAD1, PAD2, PAD3, PAD5/7 132

Figure 5.1 2 A Northern blot analysis of the PAD array mRNA transcripts 135

Figure 5.2 1 Construction of pHD451-Ty (XhoI) 139

Figure 5.2 2 Localisation of an ectopically expressed epitope tagged copy of PAD1 142

Figure 5.2 3 Designing an anti-peptide antibody specific to PAD1 145

Figure 5.2 4 The PAD1 protein is localised to the cell surface membrane 147

.Figure 5.3 1 The effect of denaturation temperature upon PAD1-Ty detection by Western blot. 150

Figure 5.4 1 PAD1 is a stumpy form specific protein 154

Figure 5.4 2 Slender and stumpy form populations can be distinguished by flow cytometry 156

Figure 5.5 1 Designing the antigenic peptides for the α-PAD2 antibodies 159

Figure 5.5 2 The lifecycle protein expression of PAD2 161

Figure 5.6 1 Designing the antigenic peptides for the α-array antibodies 164

Figure 5.6 2 The expression of PAD proteins though the T. brucei lifecycle 167

Figure 5.6 3 Identification of bands detected by the α-PAD array generic antibody 169

Figure 5.7 1 Expression of PAD1 through differentiation to the procyclic form. 171

Figure 5.8 1 Expression of PAD2 through differentiation 173

Figure 5.9 1 The expression of PAD proteins through differentiation 176

Figure 5.10 1 The correlation between PAD1 expression and EP expression during differentiation 179

Figure 5.11 1 PAD expression under differentiation conditions 182

Figure 5.12 1 The kinetics of cold-shock 185

Figure 5.13 1 The affect of varying the temperature of cold shock and the degree of parasitaemia upon PAD1 expression 187

Figure 5.14 1 The effect of cold shock of stumpy form cells upon PAD1 expression 190

Figure 5.14 2 The effect upon PAD1 expression of varying the temperature of cold-shock and the degree of parasitaemia 192

Figure 5.14 3 The effect of cold shock and FCS deprivation upon PAD1 protein expression 194

Figure 5.14 4 The affect of cold shock and pH upon PAD1 protein expression in stumpy form trypanosomes 196

Figure 5.15 1 The effect of cold shock upon PAD family protein expression 198

Figure 5.15 2 The affect of foetal calf serum deprivation and cold shock upon PAD array protein expression 200

Figure 5.15 3 The affect of pH and cold shock upon PAD array protein expression 202

Figure 5.16 1 The affect of cold shock upon the expression of PAD2 protein 204

Figure 5.17 1 PAD2 localisation in stumpy forms at 37°C 207

Figure 5.17 2 PAD2 localisation in cold shocked stumpy form cells 209

Figure 5.18 1 PAD2 in procyclic form cells 211

Figure 6.1 1 The effect of ectopic expression of PAD1 in combination with cold shock upon the sensitivity to cis-aconitate 218

Figure 6.1 2 The effect of incubating cis-aconitate with cells ectopically expressing PAD1 during cold shock 220

Figure 6.2 1 Making PAD1:pALC14 RNAi DNA construct 224

Figure 6.2 2 The effect of targeting PAD1 by RNAi for differentiation in monomorphic cells 226

Figure 6.3 1 The effect of ablating the mRNA transcripts of the entire PAD array 229

Figure 6.4 1 Design of the PAD1 knockout constructs and the differentiation capacity of a cell line containing them 232

Figure 6.4 2 Rescue of PAD1 knockout cell lines 235

Figure 6.4 3 A PCR showing that PAD1, hygromycin resistance and neomycin resistance genes were still present in the PAD1 'null' mutants 238

Figure 6.6 1 Transient RNAi in pleomorphic trypanosomes 243

Figure 6.6 2 The effect of PAD family RNAi upon differentiation in morphologically pleomorphic cells 246

Figure 6.6 3 Transient RNAi on slender AnTat 1.1 cells. Slender form AnTat 1.1 cells were transiently transfected with an RNAi construct to ablate all the PAD family transcripts 250

Figure 6.6 4 The effect of stably inducing RNAi upon pleomorphic cells 254

Figure 6.6 5 The effect of stable PAD RNAi upon the differentiation of cells in response to cis-aconitate 256

Figure 6.6 6 The effect of PAD ablation upon cold shock induced EP procyclin expression.....258

Figure 6.7 1 The ability of PAD RNAi and parental cells to respond to Pronase.....260

Figure 6.8 1 Transport of 14C citrate in Xenopus laevis oocytes micro-injected with PAD1 synthetic mRNA 264

Figure 7.7 1 A model of citrate/ *cis*-aconitate mediated differentiation.....282

List of Tables

Table 3.5 1 A summary of the PAD protein bioinformatics data.....99

Table 3.8 1 A summary of the characterised hits from a PSI-BLAST using PAD1 as a query sequence 105

Declaration

I hereby declare that all of the work presented in this study is my own, unless otherwise acknowledged. Furthermore, this thesis was written by me in its entirety.

Acknowledgements

No one completes a PhD without a vast amount of help from the people around them. I would like to take the opportunity to acknowledge this, and to thank them.

I'd like to thank all members of the Matthews lab, past and present, for their invaluable advice and support. In particular, I would like to thank: Deborah Hall, for all of her help and hard work; Katelyn Fenn, for helpful suggestions with the Bioinformatics chapter; and Pegine Walrad, for medicinal visits to the Edinburgh Whisky Society.

I'd like to thank Chris Brown for dealing patiently with endless requests for references.

I'd like to thank Rosa Marchetti and Kiaran Kirk for all of their help with the citrate transport assays. Additionally I'd like to thank the Company of Biologists for awarding me their travel fellowship that allowed me visit the Kirk lab.

I'd like to thank Keith Matthews for his scientific mentorship, his unswerving support and his investment both in me and this study. I would like to especially thank him for keeping his faith (or at least pretending to) when nothing was going as planned.

I'd thank Athina Paterou for always listening, taking care of me and for never letting me forget that she believes in me. I wouldn't have got through it all without her and I am profoundly grateful.

I'd like to thank my Grandpa, Professor Jack Dean, for always being there, and his help throughout my life. I'd like to thank my brother, Euan Dean, for regular lessons in humility while we were growing up.

Finally, I'd like to thank Alison Dean for all of her love and support, and for rising to the challenge of raising two stubborn and argumentative boys.

Abbreviations

aa	Amino acid
bp	Base pair
BLAST	Basic local alignment search tool
CAP5.5	Cytoskeleton associated protein 5.5
COX	Cytochrome oxidase complex
DAPI	4,6-diamidine-2, phenylindole
DIG	Digitoxin
DFMO	DL-α-difluoromethylornithine
DMSO	Dimethyl sulfoxide
DiD1	Defective in differentiation cell line 1
DIG	digoxygenin
DNA	Deoxyribonucleic acid
EATRO	East African Trypanosomiasis Reseacrch Organisation
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EP	Glutamic acid proline procyclin
ESAG	Expression site associated gene
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FITC	Fluorescein isothiocyanate
g	Gravities
GPEET	Glycine–proline–glutamic acid–glutamic acid–threonine procyclin
GPI	glycosylphosphatidylinositol
GPI-PLC	glycosylphosphatidylinositol-specific phospholipase C
HRP	Horseraddish peroxidase
K	Thousand
kb	Kilobase-pair
kDa	Kilodaltons
kDNA	Kinetoplast DNA
LB	Luria Broth
KO	Knockout
mm	Millimeter
mM	Millimolar
MOPS	3-(N-morpholino)-propanesulfonic acid
mRNA	Messenger RNA
PAD	Protein(s) associated with differentiation
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PSI-BLAST	Position-specific BLAST
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolutions *per* minute
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIF	Stumpy induction factor
SMB	Single marker bloodstream
TBS	Tris-buffered saline
TAE	Tris-acetate EDTA
TetR	Tetracycline repressor protein
TRITC	Tetramethylrhodamine isothiocyanate
UTR	Untranslated region
VSG	Variable surface glycoprotein
WHO	World health organisation
WT	Wild type
ZFPM	Zimmerman postfusion buffer
ZFPMG	Zimmerman postfusion buffer with glucose

Chapter 1

Introduction

1 Introduction

In addition to their considerable biomedical importance, trypanosomes are intrinsically interesting organisms to study. With unusual features such as complete reliance on post-transcriptional mechanisms for genetic regulation, RNA polymerase I mediated transcription of protein coding genes, antigenic variation and RNA editing, trypanosomes are subject to intensive research. Moreover, with the maturity of molecular interrogation techniques such as RNA interference and ectopic gene expression, researchers are in a position to address many of the biological questions that they can ask.

1.1 Evolution of Trypanosoma brucei spp.

The genus of *Trypanosoma* are single celled, parasitic organisms of the phylum *Euglenozoa*, the class *Kinetoplastida*, and the order *Trypanosomatidae*. It is proposed that the evolutionary ancestor of trypanosomes was a free living, non-parasitic organism (Vickerman (1994)) (Simpson *et al.* (2002)), that parasitized insects after their appearance 400 million years ago (Gaunt and Miles (2002)) (Engel and Grimaldi (2004)). A digenetic lifestyle may have subsequently developed, whereby a blood sucking insect was able to transfer the early parasite between vertebrate hosts (Baker (1994)) (Vickerman *et al.* (1994)) (Hamilton *et al.* (2004)): the first land based vertebrate appeared approximately 370 million years ago (Kardong (2002)), making this the earliest time point for this evolutionary step. This enabled the parasite to massively expand its host range, and in some cases revert to a monogenetic lifestyle, or even move away from insect parasitism altogether (for example, *Trypanosoma cobitis* is transmitted between fish by leaches (Lewis and Ball (1980)) (Hamilton *et al.* (2004)) (Gibson *et al.* (2005))).

The salivarian trypanosomes, constituting *T. brucei, T. evansi,* and *T. equiperdum*, are thought to have diverged from the stercocarian trypanosomes (such as *T. cruzi* and *T. lewisi*) when continental drift isolated Africa from the Americas

approximately 100 million years ago (Stevens and Gibson (1999)). The different clades subsequently evolved radically different strategies for parasitizing mammals, and in some cases humans and primates.

1.2 African Sleeping Sickness

African sleeping sickness, or human African trypanosomiasis (HAT), is caused by the parasite *Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense* and is thought to result in between 300,000 to 500,000 deaths each year (WHO (2006)) (Smith *et al.* (1998)). The ambiguity of these figures is due mostly to the difficulties associated with early infection detection, and a lack of a diagnostic infrastructure across affected areas. African sleeping sickness is described as being 'neglected' both by the media and by the pharmaceutical industry (Balasegaram (2008)), probably because it affects primarily poor regions with little economic power. *Trypanosoma brucei brucei* and *Trypanosoma congolense* cause Nagana in cattle which constitutes an enormous economic burden upon the African agriculture industry.

Sleeping sickness is found in sub-Saharan Africa in the tsetse fly belt, with foci of occurrences occurring primarily in rural, difficult to access regions, with little or no access to health services (Barrett *et al.* (2003)). The progression and type of disease depends on whether *T. b. gambiense* or *T. b. rhodesiense* initiate the infection. *T. b. gambiense* causes a chronic disease, and often maintains a parasitaemia for years prior to human host death; this parasite is responsible for >90% of all reported infections (Simarro *et al.* (2008)). *T. b. rhodesiense* causes an acute, fluctuating parasitaemia, and can be lethal within several months (Barrett *et al.* (2003)). The disease can be divided into two stages, depending on the phase of infection. The first stage involves infection of the haemolymphatic system, and causes symptoms of headaches, aching joints, itching and fever. As these symptoms are common to a number of ailments the first stage of HAT often goes undiagnosed, and therefore untreated. The second, neurological, phase occurs when the parasite crosses the blood/ brain barrier and invades the central nervous system; this is when sleeping sickness is most clearly diagnosed, and is characterised by confusion, loss of

coordination, sensory disturbances and disruption of the circadian rhythm (Buguet *et al.* (1993)). Sleeping sickness gains its name from the latter of these symptoms, and by this stage treatments are dangerous and often unsuccessful.

1.2.1 Treatments

Four drugs (see Figure 1.2-1) are currently licensed for HAT treatment that allow treatment for both stages of the disease and each of the two different disease causing sub-genuses. However, each is less than ideal due to emerging parasite resistance and toxic affects.

1.2.1.1 Suramin

Suramin was first synthesised by German chemists in 1916 and is a symmetrical polysulfonated naphthylamine derivative of urea (see Figure 1.2-1) and is trypanocidal against all types of *Trypanosoma* (Voogd *et al*. (1993)). The trypanocidal mechanism is not known, but it is thought that the high number of charges on the Suramin molecule cause inhibition of the parasitic enzyme. Suramin is 100 fold more active against blood form parasites than the insect form, possibly due to inhibition of glycolysis pathways and the high pinocytic uptake in these forms. Suramin is primarily used against the *T. b. rhodesiense* form of the disease. Due to its more acute nature, this form is more readily diagnosed in the early stages of the disease, although treatment failure was first demonstrated in the 1950s. Poor intestinal absorbance means that suramin is commonly given by intravenous injection; however, suramin does not cross the blood-brain barrier, and is consequently ineffective against the later stage of the disease. A number of side affects are associated with Suramin treatment, including neuropathy, rash, fatigue, anaemia, hyperglycaemia, and renal failure (Voogd *et al.* (1993)) (Figg *et al.* (1994)) (Bowden *et al*. (1996)). Despite resistance in the field being rare, cultured trypanosomes readily become resistant to acceptable doses of suramin, suggesting that extensive field resistance will ultimately be selected for.

1.2.1.2 Pentamidine

Pentamidine is a diamidine (see Figure 1.2-1) currently in therapeutic use against the haemolymphatic stage, primarily for the *T. b. gambiense* form of the disease. The mechanism of pathogenic uptake has been well characterised; P2 equilibrative nucleoside transporters (ENTs) (see Section 1.7.2) act to concentrate the drug from the extra-cellular environment such that a toxic concentration is achieved. Additionally, several P1 transporters (see Section 1.7.2) are also thought to contribute to this. The mechanism of trypanocidal action is largely unknown; studies in *Leishmania* have shown pentamidine accumulation at the mitochondria, and associated mitochondrial defects; additionally, treatment of blood form parasites with pentamidine has been shown to induce dyskinetoplastidy (Damper and Patton (1976)) (Shapiro and Englund (1990)) (Shapiro and Englund (1995)) (Basselin *et al.* (1998)). It is thought that the DNA binding properties of this chemical prevent proper segregation of the kinetoplast DNA, thus causing reduced virulence; however, given that a number of naturally occurring, mechanically transmitted, dyskinetoplastid trypanosomes occur in the field, and that cells selected for dyskinetoplastidy in culture are viable, the exact mechanism of toxicity is not clear. A number of pentamidine resistant cell lines have been shown to have reduced uptake of pentamidine, through mutations in, or loss of, P2 type transporters (Matovu *et al.* (2003)) (Bridges *et al.* (2007)). Other cell lines have been selected for that show normal uptake of pentamidine, but increased resistance to high intracellular concentrations (Berger *et al.* (1995)). All these cell lines appeared to have reduced virulence, suggesting that resistance in the field will not be easily selected for.

Pentamidine, as with suramin, is not well absorbed by the intestinal system, and consequently must be intramuscularly injected. Although highly trypanocidal, pentamidine is only effective against the hemolymphatic (and very early neuronal) stage of the disease as it does not easily cross the blood-brain-barrier. Additionally, toxic side affects include: hypoglycemia, hypotension, pain and abscess at the site of injection, nephrotoxicity, leucopenia, nausea and vomiting, azotemia and liver enzyme abnormalities (Sands *et al.* (1985)) (Doua and Yapo (1993)).

1.2.1.3 Melarsoprol

Melarsoprol is a melaminophenyl based arsenical compound (see Figure 1.2-1) and is the mostly widely used drug against the late stage disease; it is the only effective drug against *T. b. rhodesiense* late stage HAT, and the only affordable drug against *T. b. gambiense* late stage HAT. Despite being in use since 1949, the mechanism of action of melarsoprol is not known; the mechanism could be associated with ATP depletion (Flynn and Bowman (1974)) or thiol associations (Fairlamb *et al.* (1989)). Nonetheless, parasites exposed to melarsoprol undergo rapid lysis (Meshnick *et al.* (1978)). However, resistance to melamine based arsenical drugs is rapidly emerging in the field (Legros *et al.* (1999)) (Brun *et al.* (2001)) (Moore and Richer (2001)) (Stanghellini and Josenando (2001)), and cell lines selected for melarsoprol resistance *in vitro* have lost the P2 ENT transporter that is partially responsible to drug uptake (Carter and Fairlamb, (1993)) (Barrett and Fairlamb (1999)) (Mäser *et al.* (1999)) (Stewart *et al.* (2005)). However, it is thought that secondary routes for melarsoprol exist and melarsoprol resistant trypanosomes may have also lost the pentamidine transporter HAPT1 (Matovu *et al.* (2003)).

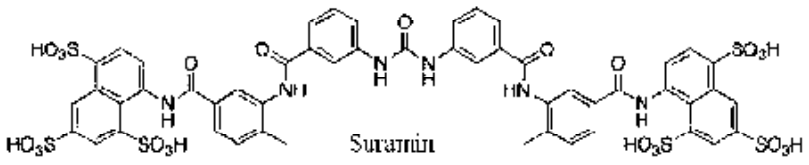
Despite its general use, melarsoprol is highly toxic to the patient, and 5 – 10% of patients suffer from reactive encephalopathy, often leading to death (Arroz (1987)) (Robertson (1963)). Moreover, even in those patients that survive the toxic affects of the drug, side effects include convulsions, headaches, itching, fever, loss of consciousness, rashes, bloody stools, nausea, and vomiting (Chappuis *et al.* (2005)). Thus, the mechanism of Melarsoprol trypanocidal activity, with a view to developing a safer alternative, is an area of research that requires urgent attention.

1.2.1.4 Eflornithine (difluoromethylornithine [DFMO]),

Eflornithine is an analogue of the amino acid ornithine that is the only licensed drug effective against the neuronal stage of *T. b. gambiense* (Burri and Brun (2003)). Eflornithine binds to the polyamine biosynthetic enzyme ornithine decarboxylase (ODC), resulting in inhibition of putrescine synthesis in the parasite (Bacchi *et al.* (1980)) (Phillips *et al.* (1987)) (Bacchi and Yarlett (1993)). The mammalian host

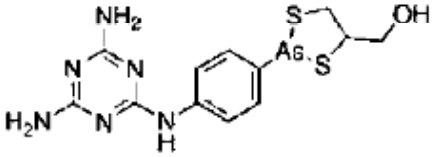
enzyme is also affected, but the parasite-specific toxicity is thought to be due to the longer half-life of the *T. b. gambiense* enzyme, such that inhibited host enzymes are more quickly re-synthesised (Ghoda *et al.* (1990)). The resistance of *T. b. rhodesiense* to DFMO is thought to be due to the faster ODC turnover in this parasite (Iten *et al.* (1997)). Other mechanisms of action may include trypanothione depletion, and a consequent vulnerability to oxidative and immunological stress (Fairlamb *et al.* (1987)). Uptake of eflornithine is thought to be transporter-mediated due to its polar nature, saturable uptake, and that resistance to eflornithine is associated with reduced uptake in procyclic forms (Bellofatto *et al.* (1987)) (Phillips and Wang (1987)). However, the transporter responsible has not yet been identified.

The pharmacological properties of eflornithine are less than ideal; low absorption kinetics means that the drug must be administered in high doses by intravenous injection. However, the high rate of mammalian ODC turnover means that this drug is relatively non-toxic in comparison with melarsoprol, and importantly, eflornithine is able to cross the blood-brain barrier, possibly though the action of an amino acid transporter or passive diffusion (O'Kane *et al.* (2006)) (de Koning (2001)). However, despite being safer than Melarsoprol, eflornithine was nonetheless reported to be fatal in 0.8% of cases (Chappuis *et al.* (2005)), and causes many of the side effects associated with Melarsoprol, albeit to a lesser intensity. Side effects include: fever, bleeding, weakness, diarrhea, nausea, stomach pain, and vomiting. Clearly, research into new treatments is of primary importance.

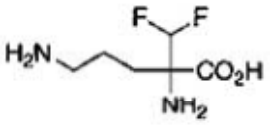


Suramin

Pentamidine



Melarsoprol



Eflornithine

Figure 1.2-1 The structure of anti- HAT drugs. Four drugs are currently licensed for use in African Sleeping Sickness treatment. There is currently a great need for new, non-toxic, drugs that are effective against the late stage disease.

1.3 Lifecycle

African trypanosomes have a complex lifecycle that exhibits highly regulated morphological, metabolic and cellular adaptations to meet the challenges of the host and the vector. Establishment and maintenance of infection, avoidance of the immune system in the mammalian host, and subsequent transmission to the tsetse fly vector requires two distinct lifecycle stages. In contrast, there are at least four distinct forms in the tsetse fly that are necessary in order to establish tsetse infection in the midgut, migrate through a complex and highly specific route to the salivary glands, undergo genetic exchange, and complete the lifecycle by infection of a mammalian host.

1.3.1 Bloodstream forms

The blood of the mammalian host presents a number of challenges to the parasite. As extra-cellular pathogens, trypanosomes have evolved a well studied paradigm of antigenic variation (Borst *et al.* (1980)) (Hoeijmakers *et al.* (1980)) (Borst *et al.* (1981)) (Pays *et al.* (1981)) (Bernards *et al.* (1984)) (Cross (1990)) (Cross (1996)) (Barry (1997)) (Marcello and Barry (2007)) (Lythgoe *et al.* (2007)) in order to evade the immune system, whereby a single variant surface glycoprotein (VSG) gene is expressed on the cell surface of the blood form parasite in very large numbers (Muñoz-Jordán *et al.* (1996)) (Navarro and Gull (2001)), acting to shield the invariant surface glycoproteins from antibody detection (Overath *et al.* (1994)) (Cross (1996)). Antibodies that bind to the VSG are sorted to the flagellar pocket of the trypanosome by hydrodynamic flow over the cell surface (Engstler *et al*. (2007)), and removed by endocytosis; this is particularly effective in stumpy form parasites, and may not be effective at preventing antibody mediated lysis at higher antibody titres. As the infection progresses, the VSG isotype is 'switched' such that a different, unrecognised VSG is expressed. In addition, the dense layer of VSG molecules prevent access to the cell surface membrane by components of the alternative complement pathway, thus providing protection against complement mediated lysis (Ferrante and Allison (1983)) (McLintock *et al* (1993)).

Blood form parasites are trypomastigotes: the kinetoplast is positioned between the nucleus and the posterior of the cell. As with all lifecycle stages, the flagellum extends from the kinetoplast and the connected basal body to the anterior of the cell in a corkscrew-like pattern (Vaughan and Gull (2003)).

1.3.1.1 Slender forms

Upon infection of the mammalian host, the metacyclic form differentiates to the slender form (see Figure 1.3-1), re-enters the cell cycle, and expresses a VSG antigen from a bloodstream-specific active expression site (Nagoshi *et al.* 1995)) (Borst *et al.* (1998)) (Cross *et al.* (1998)) (Borst and Ulbert (2001)). The blood of the mammal provides a high extra-cellular concentration of glucose which the trypanosomes utilise as a carbon and energy source. All energy generation in the blood forms is provided by glycolysis in the glycosomes and, as such, the single mitochondrion is simple and tubular, with no cristae or elaborations (Vickerman *et al.* (1988)). The pyruvate end-product is excreted. Stains for mitochondrial enzyme activity are negative on the slender mitochondrion (Vickerman (1965)) (Brown *et al.* (1973)) (Hamm *et al.* (1990)) (Bass and Wang (1991)) (Vassella and Boshart (1996)), and antibodies against mitochondrial enzymes show that these proteins are not significantly present (Tyler *et al.* (1997)).

The slender forms are approximately 25μm long and 5μm wide (Robertson (1913)). The kinetoplast is positioned approximately equidistant between the nucleus and the posterior of the cell, and the flagellum extend from this point past the anterior of the cell, where it becomes free and gives the parasite directional motility.

Perhaps most importantly, the slender form trypanosome is proliferative and forms the infection in the mammalian host. However, there is debate as to whether slender forms are able to complete the trypanosome lifecycle without first differentiating to the stumpy form (see Section 1.6 for more detail on this debate).

Slender forms are quickly killed by proteolytic stress (Sbicego *et al.* (1999)) and are unable to express procyclin on their cell surface in response to a cold shock of 20°C *in vitro* (Engstler *et al.* (2004)).

1.3.1.2 Stumpy forms

Slender forms differentiate to the stumpy forms (see Figure 1.3-1) in response to a high parasitaemia (Reuner *et al*. (1997)). The transforming factor, termed stumpy induction factor (SIF), has not yet been identified, but has been characterised as a soluble, heat stable, low molecular mass molecule that is secreted into the blood by the parasite, and may operate through a cAMP associated signalling pathway (Vassella *et al*. (1997)) (see 1.4.1 for further discussion).

Stumpy forms have a shortened morphology of approximately 15μm and are broader than slender forms (Robertson (1913)) (Vickerman (1985)). Additionally, stumpy forms do not have a free, anterior flagellum, are less motile than the slender form, and the nucleus often re-positions such that the kinetoplast is posteronuclear, although still in the trypomastigote arrangement (Vickerman (1985)).

The stumpy form mitochondrion is active as a pre-adaptation to life in the tsetse, and has a more elaborated structure than the slender form (Vickerman (1965)) (Vickerman (1988)). Moreover, the stumpy form mitochondrion stains positive for mitochondrial activity (Vickerman (1965)) (Brown *et al.* (1973)) (Hamm *et al.* (1990)) (McLintock *et al.* (1993) (Bass and Wang (1991)) (Vassella and Boshart (1996)) (Vassella *et al.* (1997)) and immunofluorescence microscopy shows that dihydrolipoamide dehydrogenase DHLADH is present (Tyler *et al.* (1997)).

Stumpy forms serve two crucial functions in the parasite lifecycle. Firstly, stumpy form cells are cell cycle arrested in G1/0 (Shapiro *et al.* (1984)) (Czichos *et al.* (1986)) (Ziegelbauer *et al.* (1990)) (Pays *et al.* (1993)) (Matthews and Gull (1994)) which has the consequence of limiting the parasitaemia and prolonging the life of the host, and thus increasing the chance of transmission to the tsetse. Differentiation to the arrested stumpy forms, and subsequent clearance of the prevalent VSG isotype in

the parasitaemia gives rise to the characteristic 'waves' of parasitaemia associated with *Trypanosoma brucei spp.* infection (Lythgoe *et al.* (2007)). Secondly, it has been proposed that stumpy forms and not slender forms are capable of differentiation to the procyclic form and thus transmission to the tsetse fly (Robertson (1913)) (Wijers and Willett (1960)) (Vickerman (1965)) (Vickerman (1985)) (Matthews and Gull (1994)) (Tasker *et al.* (2000)). Stumpy forms are terminally differentiated unless they progress to the procyclic form, and do not appear to be able to revert to the slender form (Vickerman (1985)) (Matthews and Gull (1994)).

Analysis of the slender to stumpy transition, and the precise ordering of events, has been complicated due to the presence of intermediate forms within a natural infection. Intermediate forms appear to have many of the characteristics of both slender and stumpy forms, and are probably competent to progress to the insect stage, but without an unambiguous marker for 'stumpiness' it is difficult to designate an individual parasite as slender, intermediate or stumpy.

1.3.2 Insect forms

The insect represents a radically different nutritional environment to the bloodstream. Additionally, the parasite must overcome the tsetse immune system, traverse from the insect midgut, across the peritrophic membrane and swim towards the anterior midgut, whereupon it re-crosses the peritrophic membrane at the proventriculus, and swims through the labrum and the hypopharynx into the salivary glands (see Figure 1.3-1) (Vickerman (1988)). Once at the salivary glands, the cell must undergo genetic exchange prior to attaching to the salivary gland and differentiating to a form pre-adapted to life in the mammalian immune system. This journey necessitates a number of proliferative stages and cell cycle arrested stages. Much about this journey is not known, such as how the parasite is able to navigate the tsetse anatomy, the triggers for differentiation at each stage, and the significance of lifecycle stage differences that have been identified. Additionally, although it has been conclusively shown that trypanosomes exchange genetic material (Sternberg *et al.* (1988)) (Gibson (2001)) (Gibson *et al*. (2006)) (Gibson *et al*. (2008)), trypanosome sex has not been observed and there is much speculation as to the exact mechanism of DNA transfer.

1.3.3 Stumpy to Procyclic form differentiation

Blood form trypanosomes are taken up into the crop, and then the midgut of the tsetse fly vector as it feeds on the mammalian host (see Figure 1.3-1) (Vickerman (1988)). Stumpy form trypanosomes subsequently initiate differentiation to the procyclic form primarily in the posterior midgut (Robertson (1913)) (Turner *et al*. (1988)) in response to a number of environmental cues (see Section 1.4 for details), and slender forms are either lysed by tsetse lectins and proteases (Sbicego *et al.* (1999)), or differentiate to the stumpy form and then the procyclic form. The differentiation events of this transition are well characterised because it has been possible to induce the synchronous differentiation of stumpy forms cells to procyclic form cells with high efficiency (Brun and Schönenberger (1981)) (Ziegelbauer *et al.* (1990)) (Matthews and Gull (1994)). Thus, a number of morphological, metabolic and surface coat changes have been described as part of a strictly ordered series of events.

1.3.3.1 Surface coat remodelling

The VSG homodimer is shed from the surface of the stumpy form parasite through the combined effect of glycosylphosphatidylinositol (GPI) -anchor hydrolysis and VSG homodimer proteolysis (Ziegelbauer *et al.* (1993)) (Gruszynski *et al*. (2006)) (Grandgenett (2007)). The GPI hydrolysis enzyme, glycosylphosphatidylinositol-specific phospholipase C (GPI-PLC) is expressed in blood forms but not procyclic forms (Bülow and Overath (1985)) (Carrington *et al.* (1989)) (Mensa-Wilmot *et al.* (1990)) (Webb *et al.* (2005)), present on the extracellular side of the cell surface membrane in stumpy forms (Gruszynski *et al.* (2003)) but not monomorphic slender forms (Bülow *et al.* (1989)), and contributes to the initial release of VSG upon initiation of differentiation. This enzyme is subsequently down-regulated as differentiation progresses, after which proteolysis of the VSG molecule by the zinc specific metalloprotease MSP-B is thought to complete VSG shedding (Gruszynski *et al*. (2006)). VSG shedding is complete within 8 hours of the initiation of differentiation (Gruszynski *et al.* (2003)) and as a consequence, endocytosis is

reduced in the procyclic form by a factor of 10 (Natesan *et al* (2007)) because the VSG surface molecule no longer recycles through the endocytic apparatus.

Within 2 hours of exposure to *cis*-aconitate, stumpy form parasites express EP procyclins on their cell surface (Ziegelbauer *et al.* (1990)) (Rolin *et al.* (1993)) (Matthews and Gull (1994)) (Vassella *et al.* (1997)), and within 6 – 12 hours GPEET procyclin is detectable (Vassella *et al.* (2000)). Similarly to VSG, these procyclins are bound to the cell surface membrane by a GPI-anchor; however, procyclins do not form homodimers and this GPI anchor is distinct from the blood form GPI-anchor in that it is insensitive to GPI-PLC, and is modified by the addition of different carbohydrate side chains such as polylactosamine (Ferguson *et al.* (1993)) (Treumann *et al.* (1997)) (Mehlert *et al.* (1998)). The function of these surface proteins are unknown. However, GPEET is down-regulated in late procyclic forms, suggesting a role in the establishment of the initial infection, and the amino acid repeats at the C terminus may provide protection against digestive enzymes in the tsetse midgut (Acosta-Serrano *et al*. (2000)). Moreover, despite EP procyclin null mutants showing no phenotype in culture, a reduction in the establishment of heavy infections was observed (Ruepp *et al.* (1997)).

1.3.3.2 Metabolic changes

In contrast to the glucose rich mammalian blood stream, the tsetse fly represents a glucose depleted environment for the trypanosome cell (Vickerman *et al.* (1985)) (Vickerman *et al.* (1988)). Thus, the procyclic cell develops an elaborated mitochondrion with an overall volume increase from 5% to 25% of cellular volume, exhibiting a branched network of discoid cristae (Vickerman *et al.* (1988)). Moreover, the procyclic form is enriched for cytochrome oxidase subunits associated with the electron transport chain (Tasker *et al*. (2001)) (Mayho *et al*. (2006)), and switches metabolism from glycolysis to utilising the amino acid proline as a carbon source *in vivo* (Evans and Brown (1972)) (Lamour *et al.* (2005)) (Bringaud *et al.* (2006)). However, although the glycosomes change shape from spherical to rod-shaped (Vickerman (1988)), *in vitro* they are still functional and the procyclic form

cell preferentially utilises glycolysis for its energy requirements when glucose is available (Kuile and Opperdoes (1992)) (Lamour *et al.* (2005)).

1.3.3.3 Morphological changes

Differentiation to the procyclic form is accompanied by a number of morphological changes. The trypanosome increases in length from approximately 15μm to 25μm, with a specific elongation of the post-kinetoplastic portion of the cell body (Vickerman (1985)) (Matthews *et al.* (1995)), resulting in the repositioning of the kinetoplast away from the posterior of the cell. Additionally, the flagellum extends beyond the anterior of the cell, contributing to the increased length of the procyclic form. The resulting procyclic form is long and slim, and re-enters the cell cycle to proliferate in the tsetse midgut.

1.3.4 Mesocyclic form

Procyclic form trypanosomes differentiate to the mesocyclic form in the tsetse fly anterior midgut (Van den Abbeele *et al.* (1999)) (see Figure 1.3-1). *In vivo* morphological observations demonstrate an increase in length from approximately 25μm to 34μm in the anterior midgut (Van den Abbeele *et al.* (1999)). Interestingly, mesocyclic trypanosomes in the anterior midgut appear to be cell cycle arrested in G1, leading us to draw parallels with the stumpy form that is receptive to a differentiation trigger.

The mesocyclic trypanosome is less well characterised than the blood forms and the procyclic forms because it has not yet been possible to culture them. Additionally, the triggers that induce differentiation from the procyclic to the mesocyclic form have not yet been identified, either *in vivo* or *in vitro*, thus analysis of this life cycle stage has been restricted to morphology, karyotype and site of tsetse infection location.

1.3.5 Epimastigote form

Differentiation to the epimastigote form is accompanied by migration through the peritrophic membrane and then the proventriculus, along the foregut to the salivary

gland (see Figure 1.3-1). During this migration, mesocyclic form trypanosomes differentiate to the long post-mesocyclic trypomastigote, enter G2 arrest and differentiate to the short epimastigote form (see Figure 1.3-1) (Vickerman (1988)) (Van den Abbeele *et al.* (1999)) (Gibson and Bailey (2003)).

Most distinctively, the trypomastigote post-mesocyclic cell repositions the kinetoplast from midway between the nucleus and the posterior of the cell, to anterior of the nucleus. Sharma *et al*. (Sharma *et al*. (2008)) report that this repositioning is due to a decrease in the distance between the nucleus and the posterior of the cell, and that position of the kinetoplast relative to the posterior remains constant.

The differentiation to the epimastigote form is accompanied by an asymmetric cell division that results in one long and one short epimastigote form (Lewis and Langridge (1947)) (Van den Abbeele *et al.* (1999)) (Sharma *et al*. (2008)). It is thought that the long form is terminal, and that the short form attaches to the salivary glands and continues the lifecycle (Sharma *et al*. (2008)).

The short epimastigote form is attached to the salivary gland by junctional complexes between elaborations on the flagellum and the epithelial microvilli membrane (Vickerman (1985)). Although DNA exchange has yet to be visualised, studies using fluorescent proteins suggest that sexual reproduction occurs at high efficiency in epimastigotes prior to attachment to the salivary gland (Gibson *et al*. (2008)).

The epimastigote form expresses the cell surface protein family BARP (brucei alanine rich proteins). Although originally identified as a blood form surface molecule (Nolan *et al.* (2000)), BARP was recently shown to be specific for the epimastigote form (Urwyler *et al.* (2007)), and will provide a valuable diagnostic tool in dissecting out the molecular regulatory mechanisms of this lifecycle stage. Procyclin mRNAs are also expressed in epimastigotes, but protein was not detected either by immunofluorescence (Urwyler *et al*. (2005)) or MALDI-TOF (Acosta-Serrano *et al*. (2001)), and none as yet have a clearly assigned function.

1.3.6 Metacyclic form

Short epimastigotes differentiate to the metacyclic form while still attached to the insect salivary gland (Vickerman (1985)), and enter G1 cell cycle arrest in readiness for completing the lifecycle (see Figure 1.3-1). Metacyclic form trypanosomes are pre-adapted to evade the mammalian immune system through the expression of a variable surface antigen. There are a number of interesting differences between metacyclic VSG expression and blood form VSG expression. Metacyclic cells express a more limited VSG repertoire (less than 27 in one stock of *T.b. rhodesiense*, compared to approximately 1000 in blood forms (Turner *et al.* (1988))), do not undergo antigenic variation, and do not re-arrange their telomeric DNA in order to activate VSG expression (Lenardo *et al*. (1986)) (Graham *et al.* (1990)) (Matthews *et al.* (1990)). Thus, the same 15 – 20 metacyclic VSGs appear to be expressed when *T. brucei* is cyclically passaged through tsetse. Additionally, metacyclic VSG expression is mono-cistronic with no expression site associated genes (ESAGs) between the promoter and the VSG (Alarcon *et al*. (1994)) and subject to control exclusively at the level of transcription initiation (Graham and Barry (1995)) by a locus specific mechanism (Ginger *et al*. (2002)). Metacyclic VSG expression is heterogeneous within the population (Le Ray *et al.* (1978)) (Barry *et al.* (1979)) (Hajduk and Vickerman (1981)) (although only 1 metacyclic VSG isotype is expressed at the cell surface membrane of each cell), and the metacyclic VSG expression sites are found in the telomeres of large chromosomes, in contrast to the majority of blood form VSG expression sites that are found mostly in chromosomes less than 1000kb (Cornelissen *et al*. (1985)) (Lenardo *et al*. (1984)). Upon infecting the mammalian blood, the metacyclic VSG is expressed for approximately 6 days prior to switching to a blood form expression site (Esser *et al.* (1982)).

During the differentiation process generating metacyclic forms, the glycosomes revert from rod shaped to the characteristic spherical shape observed in blood forms, and the mitochondrion reduces from a branched to a simple structure (Vickerman (1988)), suggesting a metabolic pre-adaptation to the mammalian blood stream. In addition, the flagellar outgrowths that promote attachment to the salivary gland

epithelial cells are reduced and the cell is subsequently released into the tsetse salivary gland lumen such that upon tsetse feeding the metacyclic is injected into the blood stream (Tetley and Vickerman (1985)).

Perhaps most strikingly, during this transition the kinetoplast repositions to a trypomastigote arrangement, whereby the kinetoplast is located between the nucleus and the posterior of the cell (Vickerman *et al.* (1988)). The function of the repositioning of the kinetoplast throughout the trypanosome lifecycle is unknown; one might speculate that repositioning would effect the function of the flagellum throughout the lifecycle, or possibly be associated with different modes of metabolism. The trypanosome goes to great lengths to regulate the relative position of the kinetoplast suggesting it has important implications.

It has been noted in numerous studies that trypanosomes do not complete their lifecycle through tsetse with high efficiency (Letch (1984)) (MacLeod and Turner (1999)) (Peacock *et al* (2007)); however, once salivary glands are colonised, the tsetse remains infective until it dies, and thus have the potential to infect a high number of hosts throughout its 3 month lifespan.

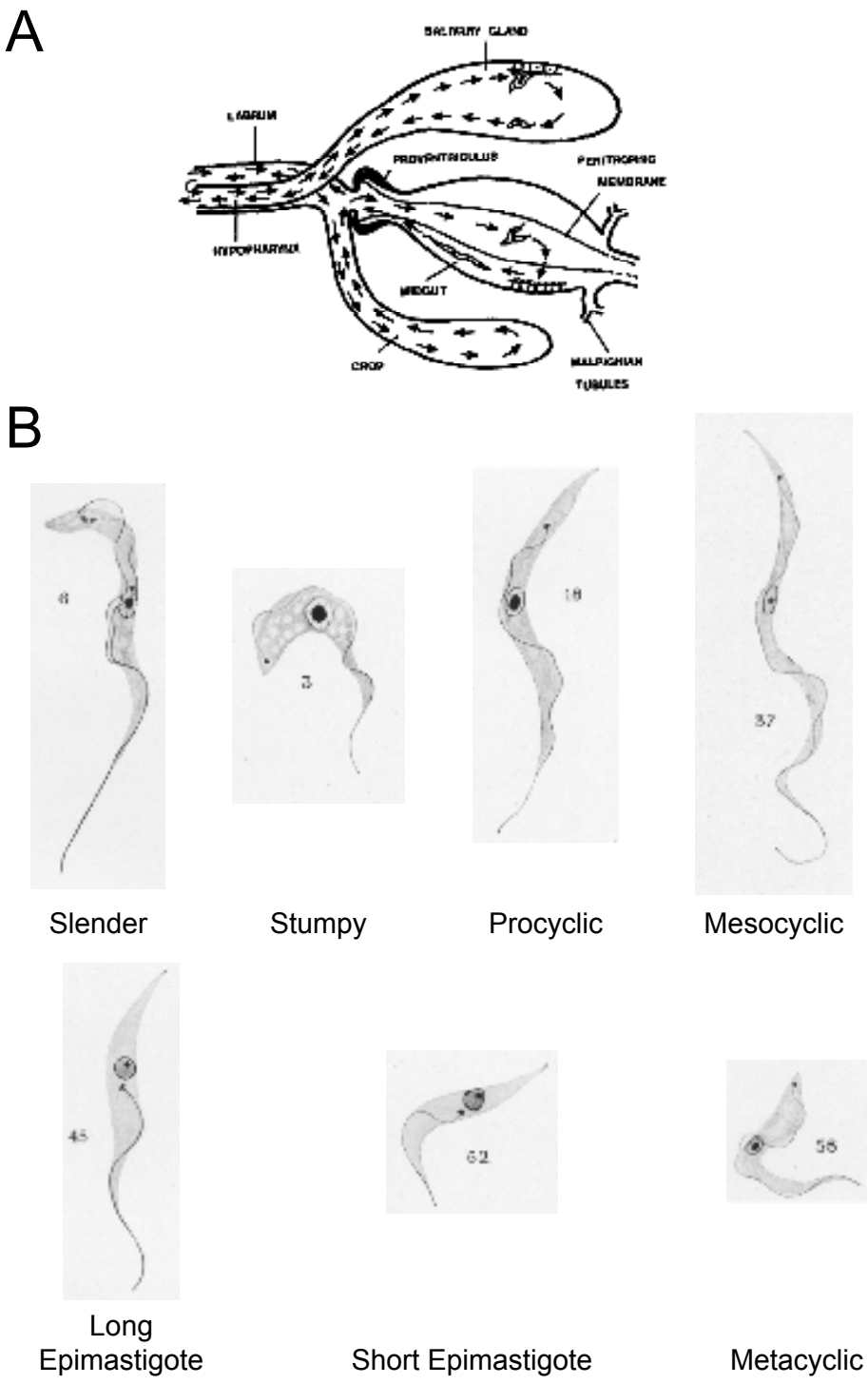


Figure 1.3-1 The trypanosome lifecycle (A) The tsetse intestinal system is shown (adapted from Vickerman (1988)). (B) The different lifecycle stages are shown (adapted from Robertson (1913)).

1.4 Chemicals reported to induce slender to stumpy form differentiation

In an naturally occurring, pleomorphic trypanosomes infection, slender forms are thought to differentiate to the stumpy form in response to cell density (Black *et al.* (1985)) (Seed and Sechelski (1989)) (Hamm *et al.* (1990)) (Reuner *et al.* (1997)); further analysis showed that this was due to an unidentified extra-cellular factor released by the trypanosomes, termed stumpy induction factor (SIF) (Vassella *et al*. (1997)). The molecular pathways that regulate and control the differentiation of slender forms to stumpy forms remain largely unidentified (Vassella *et al.* (2001)) (Pfister *et al.* (2006)), although some progress has been made in characterising the types of pathway that might be involved, chiefly through the use of membrane permeable chemicals and inhibitors. Progress in this field has been hampered by the lack of useful and unambiguous markers for the stumpy form, and hence the development of a quantitative bioassay. A number of markers do exist, but are associated with mitochondrial activation, which has been shown to be implicated in oxidative stress response and cell death in eukaryotic cells (Giffin *et al.* (1986)) (Giffin and McCann (1989)) (Lin and Beal (2006)) (Sapienza *et al.* (2008)).

1.4.1 The role of cAMP signalling

Intracellular levels of cAMP have been implicated in G1 arrest in eukaryotic cells for a number of years (Froehlich and Rachmeler (1972)), and cells of malignant origin incapable of cell cycle control have very low, or no, membrane adenyl cyclase activity (Granner *et al.* (1968)) (Makman (1970)). The first investigation of cAMP levels in *Trypanosoma spp*. showed that the stumpy form in *Trypanosoma lewisi* contains cAMP levels approximately double that of dividing, slender forms from an early parasitaemia (Strickler and Patton (1975)). This was extended to include the intracellular concentrations of cAMP at different stages of parasitaemia in *Trypanosoma brucei* (Mancini and Patton (1981)) (Vassella *et al.* (1997)) and in response to conditioned media containing SIF (Vassella *et al.* (1997)). Further, several chemical treatment affected stumpy formation, whereby a cAMP

phosphodiesterase (PDE) inhibitor, theophylline, blocked stumpy formation (Reed *et al.* (1985)), and a different PDE inhibitor, etazolate, induced stumpy formation (Vassella *et al.* (1997)). However, etazolate has other possible targets in the cell, such as chloride channels.

The most compelling evidence that cAMP signaling plays an important role in differentiation to the stumpy form was the observation that a membrane cAMP permeable derivative was able to induce transformation of slender cells to stumpy form cells (Vassella *et al*. (1997)). The authors treated pleomorphic cells with 8-(4-chlorophenylthio)-cAMP (or pCPTcAMP), a membrane-permeable derivative of cAMP that is relatively resistant to degradation by phosphodiesterases, and demonstrated differentiation to the stumpy form in response. Non membrane-permeable cAMP derivatives and 8-(4-chlorophenylthio)-cGMP (or pCPTcGMP) had no effect, suggesting intracellular cAMP signaling. Interestingly, despite monomorphic cells being unable to respond to SIF in conditioned media, either though stumpy formation or increases in intracellular cAMP concentrations, monomorphic cells can differentiate to stumpy forms in response to pCPTcAMP, suggesting that they are not defective in all parts of the stumpy formation pathway (Breidbach *et al.* (2002)).

Further investigation of the role of cAMP signaling in suggested that it was not the cAMP analogue itself that stimulated differentiation to the stumpy form, but rather the 5'-AMP/adenosine analog products of cAMP hydrolysis (Laxman *et al*. (2006)). The authors showed that non-hydrolysable, membrane-permeable, derivatives of cAMP did not induce stumpy formation, whereas membrane-permeable hydrolysis products of AMP/adenosine (such as 8-pCPT-2'-O-Me-ado) did, suggesting that AMP stimulates stumpy formation by an unknown route. The SIF induced signaling pathway of stumpy formation has not yet been characterised, and few genes involved have, as yet, been identified (Vassella *et al.* (2001)) (Pfister *et al.* (2006)).

1.4.2 Thiazolidinediones

Thiazolidinediones such as troglitazone, rosiglitazone or ciglitazone are used as anti-diabetic drugs due to their effects upon fatty acid and glucose metabolism (Henry

(1996)) (Wolffenbuttel and Graal (1996)) and have been implicated in cell growth inhibition and apoptosis in rat mesangial cells (Tsuchiya *et al.* (2003)), and adipocyte differentiation of preadipocyte and mesenchymal stem cell lines (Lehmann *et al.* (1995)). This work has recently been extended to determine whether there was an effect upon *T. brucei* blood form growth and differentiation (Denninger *et al.* (2007)).

The authors report that troglitizone induces stumpy formation in a dose dependent manner, and assay this using transformation to procyclic form efficiency, growth inhibition, cell cycle position, and mitochondrial activation. However, although their extensive experiments suggest a possible mitochondrial activation that is in accordance with stumpy formation, some of the mitochondrial subunits do not show the extent of upregulation that one would expect. For example, DHLADH protein expression was only increased 1.6 fold with respect to monomorphic slender form cells, and levels were not compared with that found in 'true' stumpy forms. Perhaps more worryingly, troglitizone treated cells did not show an accumulation in G1 as judged by flow cytometry and karyotype analysis; in contrast, the troglitizone treated population of cells could not be distinguished from non-treated controls by cell cycle position. Moreover, although the authors report an enhanced differentiation to the procyclic form in the troglitizone treated cells as judged by immunofluorescence analysis of VSG loss, differentiation was not quantified, and data regarding the number of cells expressing procyclin was not shown. Furthermore, after 8 hours exposure of troglitizone treated cells to *cis*-aconitate, there did not appear to be high levels of differentiation, demonstrating that differentiation was not synchronous as would be predicted with true stumpy forms. Troglitizone treated cells did show an initial enhanced cell growth in procyclic form media, but subsequently die at 10 days post citrate/ *cis*-aconitate induction, whereas control cells continue to grow, albeit slowly. Additionally, morphological comparisons using a pleomorphic strain to determine whether troglitizone induces morphological differentiation to the stumpy form were not compared with a non-treated control. Hence the observed morphological changes could represent normal stumpy form differentiation in response to SIF (Vassella *et al*. (1997)).

In summary, it is not clear whether troglitizone induces slender to stumpy form differentiation, and further analysis needs to be done. Nonetheless, the anti-proliferative properties of troglitizone may provide a useful route to trypanosomiasis control, as drugs of this class are already in use to treat diabetes.

1.4.3 DL-α-difluoromethylornithine (DFMO)

DFMO acts as a specific and irreversible inhibitor of ornithin decarboxylase (ODC), which catalyses the conversion of ornithine to putrescine, and thus causes an intracellular polyamine deficiency. It was originally reported to have cancer cell anti-proliferative activity and is currently in development as an anti-trypanosomiasis drug against the late stage of the *T. b. gambiense* infection (Bacchi *et al.* (1980)) (Bacchi *et al.* (1994)) (Van Nieuwenhove *et al.* (1985)) (Doua *et al.* (1987)) (see Section 1.2.1.4).

DFMO was reported to cause differentiation to the short stumpy form in the late 1980s (Giffin *et al.* (1986)). (McCann *et al*. (1986)) (Giffin *et al.* (1989)). DFMO treated cells demonstrated mitochondrial activation, stained positive for NAD diaphorase activity and exhibited a stumpy morphology; however, they were unable to transform to the procyclic form when incubated in SDM79 media and did not show uniform accumulation in their cell cycle. The ability of DFMO treated cells to differentiate synchronously in response to citrate/ *cis*-aconitate was not tested. However, the high rate of cell death combined with the appearance of multinucleate cells suggests that the stumpy formation and mitochondrial activation reported may be accounted for by cellular stress induced by putrescine depletion. Additionally, stumpy forms have subsequently been shown to have similar levels of ODC activity as slender forms, suggesting that putrescine depletion does not occur in this form (Bass and Wang (1991)). Further analysis on the effect of DFMO and the role of polyamines in slender to stumpy differentiation may be required, especially in light of the therapeutic value of DFMO in treating African Trypanosomiasis.

1.5 Conditions reported to induce stumpy to procyclic form differentiation

Several blood form to procyclic form differentiation triggers have been identified, these exhibiting differing efficiencies of induction.

1.5.1 Incubation at 27°C

The first recorded stimulus that was able to trigger blood form to procyclic form differentiation *in vitro* was a temperature shift from 37°C to 27°C (Evans and Brown (1971)) (Srivastava and Bowman (1971)) (Srivastava and Bowman (1972)) (Brown *et al.* (1973) (Ghiotto *et al.* (1979)) (Bienen *et al.* (1980)). Several researchers reported growth of trypanosomes that were sensitive to cyanide when parasites harvested from blood were incubated at 27°C, demonstrating a change to the insect form of metabolism. This differentiation was shown to occur over the course of 72 to 96 hours, with subsequent stable growth. Further analysis documented the kinetics of differentiation, with approximately 20% and 90% differentiation after 24 and 48 hours respectively, depending on the effects of different media substrates (Bienen *et al*. (1980)).

1.5.2 Incubation at 27°C and the addition of tri-carboxylates

Brun and Schönenberger (Brun and Schönenberger (1981)) showed that highly efficient and synchronous differentiation to the procyclic form could be induced by the addition of the tricarboxylates Krebs cycle intermediates citrate and *cis*-aconitate. A concentration of 2 – 10 mM *cis*-aconitate was the most efficient at stimulating differentiation, followed by citrate in the same range; differentiation was not detected using 0.1 mM of either of these tri-carboxylates, was saturated at concentrations of 8 mM, and inhibited at concentrations above 16 mM, possibly due to cytoxic effects. Interestingly, and inexplicably, although no other tri-carboxylate was able to induce efficient differentiation, a mixture of all the tri-carboxylates in the absence of citrate/ *cis*-aconitate was able to induce differentiation. Nothing further has been reported on this anomaly, suggesting that it may have been an experimental artifact.

This work was extended (Czichos *et al.* (1986)) to show that, although trypanosomes would differentiate at 37°C, a temperature reduction to 27°C in addition to citrate/ *cis*-aconitate was necessary for the establishment of a dividing procyclic form population. Furthermore, the authors showed that exposure to *cis*-aconitate for >8 hours was necessary for a level of differentiation >2% after 72 hours, suggesting that these cells require a long exposure to citrate/ *cis*-aconitate to become fully committed to the procyclic form. It is worth noting, however, that this latter study utilized the monomorphic strain MITat 1.4 that does not give rise to stumpy form cells, and that stumpy form cells express procyclin and undergo the other events of differentiation within 4 hours of exposure to citrate/ *cis*-aconitate (Ziegelbauer *et al*. (1990)) (Matthews and Gull (1994)) and will differentiate at 27°C in the absence of citrate/ *cis*-aconitate. A discussion of the relative efficiencies of differentiation between monomorphic and pleomorphic cell types is detailed in Section 1.6.

The ability to efficiently and synchronously recreate the differentiation of blood forms to procyclic forms *in vitro* has been invaluable in the dissection and analysis of the events associated with this transition. However, the high concentrations required meant citrate/ *cis*-aconitate was excluded from being the physiological trigger. In 2004, however, Engstler *et al.* showed that when stumpy form trypanosomes are subjected to a cold shock of >15°C (i.e. from 37°C to 20°C), differentiation to the procyclic form was able to proceed at concentrations of *cis*-aconitate as low as 10 μM, albeit with slower kinetics. The concentration of citrate in the tsetse midgut is not known; although citrate concentrations in whole tsetse extracts has been reported to be 15.9 μM (Hunt *et al.* (1994)), this study did not determine the midgut concentration specifically. However, citrate concentrations in human blood plasma has been reported to vary between 80 μM to 230 μM (Jacobs and Lee (1964)) (Taylor *et al.* (1998)). Thus, trypanosomes taken up as part of a blood meal are exposed to concentrations of citrate that have been reported induce differentiation at 20°C, suggesting that citrate, and possibly *cis*-aconitate, may well be an *in vivo* trigger for differentiation.

1.5.3 Mild acid treatment

Rolin *et al.* (1998) investigated the role of other stimuli in promoting differentiation. The effects of mild acid treatment were investigated because firstly, mild acid treatment had been shown to play an important role in the differentiation of flagellated forms to amastigotes in *Leishmania* and *T. cruzi* (Kanbara *et al.* (1990)) (Zilberstein *et al.* (1991)) (Bates *et al.* (1992)) (Tomlinson *et al.* (1995)), and secondly, because mild acid stimulated a number of events that were associated with blood form differentiation, namely VSG release and activation of adenylate cyclase (Rolin *et al*. (1993)) (Rolin *et al* (1996)). Rolin *et al.* (1998) report that when pleomorphic trypanosomes, comprising a mixed population of slender, intermediate and stumpy form parasites, were pre-incubated at pH 5.5 prior to incubation at 27°C, differentiation to the procyclic form was induced, albeit at a lower efficiency than when cells are incubated in *cis*-aconitate. Interestingly, the authors reported that, although slender forms were quickly lysed under these conditions, intermediate forms were stimulated to differentiate to the stumpy form. This demonstrates that (a) stumpy formation is enhanced by low pH (b) stumpy like cells are more resistant to pH changes, possibly as an adaptation to tsetse colonisation. The latter was confirmed by Nolan *et al.* (2000). The mechanism of stumpy form resistance to pH change is not known, but one might speculate that it is associated with an improved ability to maintain membrane potential.

The physiological relevance of these findings is still not clear as the tsetse midgut pH exhibits an alkaline pH (Van den Abbeele *et al.* (1999)); however, the precise compartment in which trypanosomes initiate differentiation may not be the tsetse midgut, and parasites may be subjected to pH changes during uptake by the tsetse. A number of cellular factors are affected by pH change, including the operation of transporters, but the mechanism of action of mild acid-induced differentiation is not known.

1.5.4 Proteolytic stress

Differentiation of blood form trypanosomes in response to proteolytic stress was first described by Yabu and Takayanagi in 1988. The authors subjected cultured,

monomorphic *Trypanosoma brucei gambiense* to trypsin treatment and observed levels of differentiation comparable with that obtained through citrate/ *cis*-aconitate treatment; however, the percentage of cell death was high and a small founder population survived and differentiated (see Matthews and Gull (1997)), thus differentiation efficiencies cannot be usefully compared to citrate/ *cis*-aconitate induced differentiation. Their work was extended using pleomorphic trypanosomes to include a mixture of microbial proteinases that induced differentiation at a higher efficiency than citrate/ *cis*-aconitate (Hunt *et al*. (1994)) (Sbicego *et al.* (1999)), and was lytic to long slender trypanosomes (Sbicego *et al.* (1999)).

Studies using a tsetse midgut homogenate demonstrated that activities triggering differentiation were destroyed upon heating, and inhibited using trypsin inhibitors, suggesting that a protease component played an important role *in vivo* (Imbuga *et al.* (1992)). Thus, proteinase treatment of trypanosomes may mimic the tsetse midgut, where a number of tsetse digestive enzymes are thought to operate. The mechanism of operation is not known, however, and is thought to operate differently from the citrate/ *cis*-aconitate pathway.

1.5.5 Replacement of glucose with glycerol in the media

Milne *et al.* (Milne *et al.* (1998)) report the inefficient differentiation of blood form parasites at 37°C in response to replacing the glucose carbon source with glycerol. Differentiation was determined using a number of assays (including morphology, procyclin expression, CAP5.5 expression), and the resulting procyclic form trypanosomes appeared to divide with faster kinetics at 37°C than at 27°C. The authors speculated that glucose deprivation may be an *in vivo* trigger, as glucose is thought to be depleted in the blood meal within 15 minutes of tsetse feeding (Vickerman (1985)) (Mike Turner, unpublished observations).

This observation has a number of interesting implications, not least that temperature can be uncoupled from procyclic form proliferation. This is in apparent contrast to previous observations using citrate/ *cis*-aconitate as a differentiation trigger (Czichos *et al.* (1986)) (Mutomba and Wang (1995)), whereby a temperature reduction to

27°C was required in order for trypanosomes to re-enter the cell cycle. The significance of this disparity is not clear and has not been investigated further, but it does highlight that differentiation can be uncoupled from procyclic form proliferation. This is supported by the observation that glucose, although having no effect upon differentiation *per se*, was found to inhibit the growth of transforming cells (Bienen *et al.* (1980)) and established procyclic cultures (Ebikeme *et al*. (2008)). Thus, the differentiation observed by Milne *et al.* (1998) may be an example of founder effect combined with enhanced survival of procyclic form parasites in the absence of glucose.

1.6 The relationship between cell cycle position, life cycle position, and competence to differentiate

Robertson was the first person to suggest that slender cells do not transmit to the tsetse, and that intermediate and stumpy form cells form the establish the midgut infection (Robertson (1913)). Since then, this theory has been the centre of some controversy.

The principal evidence supporting the stumpy, cell cycle arrested form, as being the transmission competent form comes from experiments demonstrating synchronous differentiation of stumpy form cells (Ziegelbauer *et al.* (1990)) (Matthews and Gull (1994)). The authors showed that in a pleomorphic population, after 6 hours of *cis*-aconitate treatment only cells identified as being stumpy form were expressing procyclin (Matthews and Gull (1994)). Furthermore, an analysis of the differentiating population showed that of 175 dividing cells (as judged by karyotype), none were expressing procyclin. The authors identified 2 possible explanations for this, (a) stumpy forms only were able to initiate differentiation, or (b) the ability to respond to *cis*-aconitate depended upon cell cycle position. To investigate this further, the authors differentiated 2 monomorphic cell lines that had lost the capacity to transform into the stumpy form, and a very early pleomorphic population of cells comprising predominantly slender forms. The authors showed that these cells differentiated very poorly in the first 6 hours (<10%, compared with >95% in a stumpy population), thus verifying that stumpy cells differentiated at a

much higher frequency than slender cells in the first 6 hours. Importantly, the authors also showed that there was a delay in BrDU uptake in differentiating monomorphic cells, indicating that cells in this population enter cell cycle arrest. Thus, although the link between pleomorphic differentiation and cell cycle position is convincing, this paper does not conclusively link the ability to initiate differentiation to cell cycle position in monomorphic cells. The reduced differentiation efficiency of monomorphic slender cells could be attributed to either i) that only a small proportion of cells that are in G1 is receptive to the differentiation trigger (thus termed stumpy*) or ii) slender cells are receptive to the trigger throughout the cell cycle, but respond with slower kinetics than stumpy form cells. Attempts to address this issue by determining the cell cycle position of differentiated monomorphic cells very early through differentiation were unsuccessful (Keith Matthews, personal communication).

Opponents of this hypothesis support the view that trypanosome differentiation is distinct from differentiation in other eukaryotic cells, and reflects a simple change in gene expression levels, rather than a true differentiation (Mutomba and Wang (1996)) (Mutomba and Wang (1998)) (Li and Wang (2003)). Thus, it is argued that a specific cell cycle position or G1/0 arrest is not necessary for efficient differentiation to the procyclic form. To investigate this, a monomorphic cell line was generated that was able to inducibly ablate proteosomal subunits, and consequently enter cell cycle arrest in G1 and G2 (Li and Wang (2003)) . The authors reported that this did not inhibit the differentiation of cells to the procyclic form, thus disproving a link between cell cycle progression and differentiation. However, the cell cycle arrest reported in this paper was not convincing and was shown to have a high death rate, possibly resulting in secondary effects. Moreover, a quantitative analysis of differentiation was not performed until 60 hours post-differentiation, such that interpretation of the data may have been complicated by procyclic form outgrowth. Additionally, their null hypothesis relies on procyclin expression being dependent upon cell cycle progression, rather than cell cycle position, which has been shown not to be the case (Matthews and Gull (1994)). Perhaps more convincingly, the authors subjected a pleomorphic population of cells to chemical inhibitors of DNA α

polymerase (using aphidicolin (APH)) and DNA synthesis (using hydroxyurea (HU)), giving rise to cells arrested in G2 that appeared to differentiate (Mutomba and Wang (1996)). However, these cells did not differentiate as well as control cells and the high rate of cell death may have complicated the subsequent analysis.

The evidence indicates that stumpy forms are able to differentiate with much higher efficiency than slender forms, and that true slender forms may not be able to differentiate at all. However, when explaining the low levels of differentiation in monomorphic cells at early timepoints, it is very difficult to distinguish between low frequency of differentiation initiation (possibly due to a requirement for cell cycle arrest in G1), and the reduced speed of differentiation once it is initiated; thus, this issue is not fully resolved.

1.7 Transporters

Transporters represent a very large subset of genes in the parasitic protozoa; between 2 – 2.5% of genes in the 'tritryp' genome database are annotated as having a transporter-like function (Berriman *et al.* (2005)) (Ivens *et al*. (2005)) (El Sayed *et al*. (2005)). In addition to their intrinsic biological importance in nutrient scavenging and environmental sensing, transporters play a central role in drug intervention in trypanosomes through either (i) enabling uptake or (ii) enabling efflux of toxic drugs For example, the *T. brucei* transporter TbAT1, that is responsible for uptake of adenosine and adenine from the host blood, plays a crucial role in sensitivity to melarsoprol, as resistant trypanosomes have been shown to lack this transporter (Carter and Fairlamb (1993)) (Carter *et al*. (1995)). Other transporters, such as HAPT1 and LAPT1, appear to mediate the uptake of pentamidine and melaminophenyl arsenical drugs (Matovu *et al*. (2003)). Additionally, transporters themselves can be good candidates as drug targets; *Leishmania* glucose transporter null mutants are not viable in the blood because glucose is an essential precursor in synthesizing glucoconjugate and complex carbohydrates (Landfear (2008)). Hence, an inhibitor of this transporter may have blood form trypanocidal activity.

Thus, understanding the role of transporters in pathogenic organisms, and the mechanism of substrate transfer, is of principal importance in the rational design of anti-parasitic drugs. Currently, bioinformatics is a poor tool for predicting transporter substrate specificity and function; however, with more experimental data and better algorithms, this can be expected to change in the next few years.

1.7.1 Adenosine binding cassette (ABC) transporters

ABC transporters are ubiquitous molecules found in prokaryotic and eukaryotic cells, both at the cell surface and in organelles. They catalyse the translocation of a diverse range of substrates across the cell surface membrane, often against the concentration gradient, using the energy release when ATP is dephosphorylated to ADP (Higgins (1992)).

Bioinformatics based prediction of ABC transporters has become a valuable and reliable tool due to the well defined domains and conserved motifs in this group of proteins. They contain two nucleotide binding domains (NBDs) that are characterised by the conserved ABC signature, the Walker A and the Walker B motifs (Walker *et al.* (1982)) (Higgins *et al.* (1985)) (Higgins *et al.* (1986)). In the cystic fibrosis transmembrane conductance regulator (CFTR) channel, one NBD is associated with 'opening', and the other is associated with 'closing' the channel (Carson and Welsh (1995)) (Cotten and Welsh (1998)) (Gadsby and Nairn (1999)) (Vergani *et al*. (2003)).

Mutations in ABC genes have been associated with a number of diseases; for examples, mutations in the NBD of the CFTR gene have been linked with cystic fibrosis. Other ABC genes have been linked with drug resistance; for example, Pgp encoded by MDR1 gene in human tumors has been shown to export multiple anti-cancer drugs (Ambudkar *et al.* (1999)) (Gottesman *et al.* (2002)) (Tombline *et al.*(2004)). To date, 18 ABC transporters have been found in *Trypanosoma brucei*, 36 in *Leishmania*, and several have been associated with drug resistance (Hendrickson *et al*. (1993)) (Chow *et al*. (1993) (Ouellette *et al*. (1998)) (Katakura *et al*. (1999)) (Shahi *et al.* (2002)).

1.7.2 Equilibrative nucleoside transporters (ENT)

Trypanosomes cannot synthesis purines de novo (Hammond and Gutteridge (1984)) (Fairlamb (1989)) and consequently have efficient means for scavenging nucleosides from the extracellular environment. Thus far, 8 cell surface membrane purine transporter have been identified in *Trypanosoma brucei* (Sanchez *et al.* (2004)) (De Koning *et al.* (2005)), and all are members of the ENT family. ENT proteins are monomeric, are predicted to have 11 transmembrane helices and utilize a pre-existing chemiosmotic gradient to translocate substrates across the lipid bilayer. Despite being named after the properties of the first mammalian member of this group to be characterised, some protozoan members concentrate their substrate utilizing a proton gradient (Landfear (2001)), whereas others transport non-nucleoside substrates such as nucleobases (Henriques *et al.* (2003)).

Protozoan ENTs have been distinguished based upon their substrate specificity (Carter and Fairlamb (1993)): P1 transporters are specific for inosine and adenosine, whereas P2 transporters are able to transport adenine, and the trypanocidal melaminophenyl arsenicals and diamidines. For this reason, most research into kinetoplastid ENTs has concentrated on the P2 type transporters, such as TbAT1, HAPT1 and LAPT.

1.7.3 Major Facilitator Superfamily (MFS)

The MFS represent the largest family of transporters (Henderson and Maiden (1990)) can be found in all types of cell, and allow the uptake of nutrients, excretion of metabolic end-products and environmental sensing (Mitchel (1957)) (Grifith *et al.* (1992)) (Marger and Saier (1993)) (Essenberg *et al*. (1997)) (Pao *et al.* (1998)). They are secondary transporters able to catalyse the translocation of a wide range of low molecular mass substrates across a membrane by utilising a net chemiosmotic gradient (Paulsen *et al*. (1998)). In this way, many members of the MFS utilise a cationic gradient to transport a substrate against its concentration gradient, for example; the *E. coli* lactose permease (LacY) is a member of the MFS that translocates galactoside into the bacterial cell against its concentration gradient by utilising the proton concentration difference between the intra- and extra-cellular

environments (Matsushita *et al.* (1983)). Elegantly, this proton gradient is generated by the action of active proton pumps, and in common with other MFS transporters, translocation can be reversed if the respective gradients are reversed.

Members of the MFS are predicted to have either 6, 12 or 14 transmembrane spanning helices (TMS) (Goswitz and Brooker (1995)); it is thought the 6 TMS MFS is a product of an evolutionary duplication of an ancestral 3 TMS transporter, and that the 12 TMS MFS resulted from the 6 TMS being duplicated, with an extra 2 TMS becoming duplicated in the 14 TMS (Griffith *et al.* (1992)) (Henderson and Maiden (1990)) (Paulsen and Skurray (1993)). MFS have a conserved motif between TMS 2 and 3, and another between TMS 8 and 9 (Henderson and Maiden (1990)). There is no evidence that wild type MFS act as multimers (Sahin-Tóth *et al.* (1994)); however, mutant MFS whereby several TMS have been deleted are able to be complemented by other mutants where the respective TMS are present, showing up to 60% wild type activity (Bibi and Kaback (1991)). By far the most well studied of the MFS are those containing 12 TMS; members of the 6 and 14 TMS are fewer and less well characterised.

The hydrophobic nature of transporters, and their transitory configurations, makes them inherently difficult to crystallize, meaning that few structures have been resolved. The generation of a mutant thermostable LacY that favours one conformation facilitated the crystallization of this transporter in 2003 (Abramson *et al.* (2003)); the TMS are arranged as two groups of 6 helix bundles, whereby each of the TMS passes through the membrane such that a hydrophilic circular pore is formed. In the open conformation, side chains within the TMS are available for substrate binding deep within the pore, and subsequent translocation upon conformational change. This basic model has been applied to the rest of the MFS, although it is clear that there will be variation between transporters, especially those that contain different numbers of TMS, and much work is continuing to determine the precise mechanism of transport and which residues are involved at each stage.

Other members of the MFS include the multidrug resistance (Mdr) transporters that are characterised by their ability to export a wide range of drugs from pathogenic

microorganisms or cancer cells. These appear to form numerous, weak hydrophobic bonds between the substrate and the side chains within the pore, possibly explaining their wide substrate specificity (Neyfakh (2002)). Other transporters, with more specific substrates, have fewer, stronger and more specific hydrogen bonds with a narrower range of substrates.

A number of MFS transporters have been characterised in *T. brucei* and other kinetoplastids. In particular, the Glucose transporter (GLUT) family in *Leishmania* (Langford *et al*. (1994)) (Snapp and Landfear (1997)) (Burchmore and Landfear (1998)) (Rodriguez-Contreras and Landfear (2006)) (Rodriguez-Contreras *et al.* (2007)) is of particular importance as the energy requirements of the blood form parasite are met exclusively through glycolysis.

1.7.4 Channels

1.7.4.1 Aquaporins

Aquaporins are tetrameric, membrane bound protein channels that selectively allow the transfer of small molecules, such as water or glycerol, across the lipid bilayer. They serve crucial functions in cells, such as osmoregulation, size control, and in multicellular organisms help to maintain homeostasis of the organism. Aquaporins form tetrameric channels, whereby each monomer consists of 6 α transmembrane helices. Within the tetramer, each monomer forms a distinct pore that penetrates the lipid bilayer, allowing access for substrate molecules. The α helices are orientated in such a way that hydrophobic helices face the surrounding bilayer, and hydrophilic helices face inward to form the hydrophilic pore. Each monomer contains 2 distinct Asn-Pro-Ala (NPA) signature motifs that appear to be important for specificity in some aquaporins (Fujiyoshi *et al*. (2002)).

The principal mechanism of aquaporin selectivity is size exclusion. For example, in AQ1, a mammalian water-specific aquaporin, each monomer narrows to 2.8 Å above and below the lipid bilayer such that entry of molecules larger than H_2O into the pore is excluded (Sui *et al.* (2001)). Additionally, the NDC of AQ1 contains a pair of dipoles; in order to cross the channel each water molecule must interact with the

dipole, thus isolating it from surrounding water molecules and preventing access of H+ ions. Other aquaporins are required to maintain homeostasis, cell size or osmotic potential. In plant cells, aquaporins allow access of CO_2 for photosynthesis (Uehlein *et al.* (2003)), and in the kidney duct-epithelial cells, aquaporins are released onto the cell surface from intracellular storage vesicles in response to vasopressin binding and the association with another membrane bound protein (Kamsteeg *et al*. (2007)) (Kamsteeg *et al*. (2008)), thus increasing the water permeability of the cell surface membrane.

Kinetoplastids have several characterised aquaporins that are thought to enable the parasite to survive the differences in osmotic potential between different hosts through the lifecycle, to regulate variations in the cellular volume in different lifecycle stages, and to transport small molecules. In *T. brucei*, 3 aquaporins have been characterised that are differentially regulated throughout the lifecycle, and are able to transport water for osmoregulation, and glycerol, a waste product of glycolysis (Uzcategui *et al.* (2004)).

1.7.4.2 Ion channels

Ion channels allow the highly selective passage of ions such as Na^+, Ca^+, K^+ and Cl^- across the lipid bilayer in response to an ionic gradient (Dutzler (2004)). These molecules are required to maintain a constant intracellular pH for efficient enzyme function, nervous system signalling, and also to form the ionic gradients that are required to facilitate transport of other solutes (Kavanaugh (1998)). They contain a variable number of transmembrane helices (up to 24 helices per channel) (Saier (2000)), and the mature, functional protein is often a multimer (Nelson *et al.* (1999)) (Liapi and Wood (2005)). Ion channels have particular relevance to single celled parasites such as *Trypanosoma brucei* due to the fluctuation of extracellular pH throughout the parasite lifecycle.

In contrast to other types of transporter which are found in several transitory states throughout substrate translocation, ion channels function as ‘open’ or ‘closed’ gates (Jiang *et al*. (2002)). When in the open conformation, solute passage can occur with

very high flux (up to 10^8 ions transported *per* second), and can help to build a concentration gradient by utilising the existing concentration gradient of a co-transported molecule (Kavanaugh (1998)). Ion channels can be classified by the mechanism of gate switching, including the voltage gated and ligand gated ion channels (Saier (2000)), whereby either a change in the electrical potential difference or ligand binding respectively causes a conformational change in the channel, resulting in an open or closed gate.

A number of ion channels have been annotated in the *T. brucei* genome (Berriman *et al*. (2005)), and several have been subsequently detected by proteomic analysis (Bridges *et al.* (2008)); however few have been characterised in detail due to the inherent difficulties associated with detecting hydrophobic proteins.

1.8 DiD1 generation and identification of the Proteins Associated with Differentiation (PAD) gene family

In order to help elucidate the mechanisms underlying differentiation, a cell line was selected for that was unable to initiate differentiation to the procyclic form in response to *cis*-aconitate (Tasker *et al.* (2000)) (see Figure 1.8-1). The selection regime for this relied on two biological characteristics: firstly, the variant surface glycoprotein (VSG) protects the blood form parasite against lysis by the mammalian alternative complement lysis pathway (Ferrante and Allison (1983)) and, secondly, that blood form parasites lose their VSG coat as one of the first steps of differentiation (Gruszynski *et al.* (2003)) (see Section 1.3.3.1 for details). Thus, the monomorphic slender *T. b. rhodesiense* East African Trypanosomiasis Research Organization (EATRO) 2340 GUP2965 cell line was incubated with 6 mM *cis*-aconitate in SDM79 at 27°C and monitored for VSG loss. When 99% of cells were identified as VSG-negative by immunofluorescence analysis, the cells were pelleted and inoculated into a mouse in order for the remaining 1% of parasites to form an infection, thus providing a very powerful selection pressure for cells that do not lose their VSG surface coat in response to *cis*-aconitate. After 4 cycles of differentiation and selection, the population of cells was clearly less able to differentiate, and the Defective in Differentiation cell line 1 (DID1) was cloned for analysis.

Analysis of the DiD1 cell line defect revealed that the majority of cells were impaired in their ability to differentiate in response to *cis*-aconitate: in addition to not losing their VSG surface coat, these cells did not gain procyclin, undergo kinetoplast repositioning, mitochondrial elaboration, or any of the hallmarks of differentiation. The authors concluded that since VSG loss is an early indicator of differentiation, the DiD1 cell lines fails to initiate differentiation in response to *cis*-aconitate (Tasker *et al.* (2000)).

Importantly, a small sub-population of DiD1 cells were able to differentiate, comparable to differentiation in the absence of *cis*-aconitate (approximately 0.2% of cells can gain a procyclin coat after 72 hours incubation at 27°C regardless of the presence or absence of *cis*-aconitate). This suggested that it was not differentiation *per se* that was impaired, but rather the *cis*-aconitate response pathway. Other differentiation triggers were not reported (Tasker *et al.* (2000)), but subsequent analysis demonstrated that both DiD1 and parental cells underwent massive cell death in response to protease and mild acid treatment, as would be expected with a monomorphic slender population (Maria Tasker, unpublished observations). However, both the DiD1 cells and the parental cells were able to establish midgut infections in tsetse (Maria Tasker, unpublished observations). This demonstrates that, *in vivo*, differentiation can operate independently of citrate/ *cis*-aconitate pathways, and that DiD1 cells were able to respond to other triggers.

Interestingly, the authors were able to rescue the differentiation defect by generating DiD1 ΔGPI-PLC mutants. In order to determine whether the DiD1 cells that were able to differentiate were doing so by a non *cis*-aconitate mediated pathway, the authors removed GPI-PLC from the DiD1 cell line through sequential rounds of gene knockout, and tested this cell line to determine whether the small numbers of cells able to differentiate was now lost. The rationale for this was that it had been proposed that inefficient differentiation is facilitated by a GPI-PLC mediated pathway (Rolin *et al.* (1998)). Surprisingly, when ΔGPI-PLC DiD1 cells were grown to a very high parasitaemia in immunosuppressed mice, cells reverted to a morphologically pleomorphic population and were able to differentiate efficiently in response to *cis*-aconitate. This effect was not shown when cells were grown in

culture, and the reversion to a pleomorphic morphology was probably not a specific characteristic of DiD1 cells. The authors speculated that the rescue of the differentiation phenotype was due to the attenuated *in vivo* growth caused by GPI-PLC deletion (Webb *et al.* (1997)) allowing the accumulation of a cryptic, stumpy-specific, characteristic that allows reception of the *cis*-aconitate signal.

In order to further understand the DiD1 defect, and thus to gain insight into the mechanisms underlying *cis*-aconitate induced differentiation, Davies and Levin (University of Manchester) performed a macroarray hybridisation analysis to identify transcripts that were differentially regulated between the DiD1 and the parental cell line. A macroarray was generated from a sheared genomic DNA library (a kind gift from Prof. E. Ullu, Yale University, USA) and probed with radiolabelled cDNA from the DiD1 and the parental cell line. This identified 1 clone as differentially hybridizing to DiD1 cDNA, and not parental cDNA. This clone was subsequently shown to contain sequence from two adjacent genes on chromosome 7. These genes were part of an 8 gene cluster, and named Proteins Associated with Differentiation (PAD) 1 – 8.

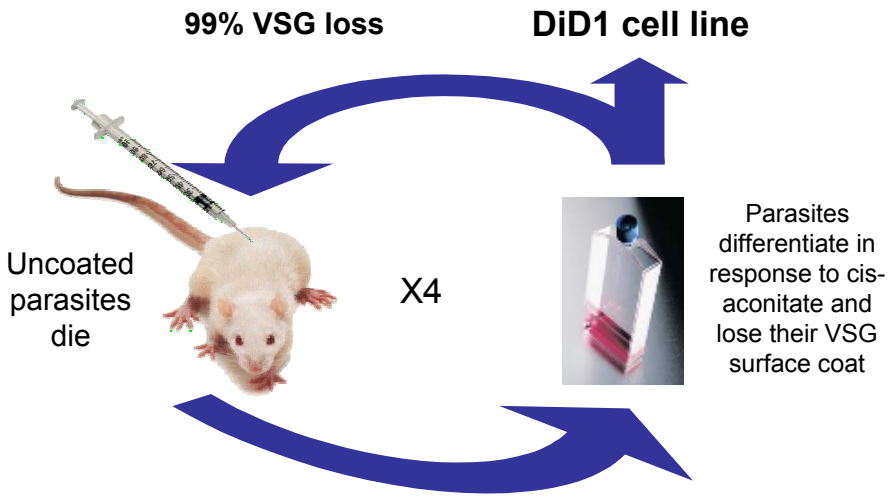


Figure 1.8-1 Selection of the DiD1 cell line. Monomorphic slender cells were induced to differentiate by incubation with 6 mM *cis*-aconitate at 27°C, and monitored for VSG loss. Subsequent to 99% VSG loss, cells were pelleted and injected into a mouse. The resulting parasitaemia was isolated and cells were purified by anion exchange prior to undergoing another round of selection. After 4 rounds of selection, the population of cells was clearly less responsive to differentiation conditions, and the Defective in Differentiation cell line 1 (DiD1) was cloned.

1.9 Aims and objectives

Work undertaken as part of this thesis has aimed to

- Determine the basis of the DiD1 defect
- Functionally characterise members of the PAD array, with particular emphasis on their putative role in differentiation.

Chapter 2

Materials and Methods

2 Materials and Methods

2.1 Trypanosomes

2.1.1 Strains and cell lines

Unless otherwise stated, the trypanosome cell lines used in this study are derived from the *Trypanosoma brucei brucei* Lister 427 strain (henceforth they are abbreviated '427'). These are laboratory adapted, monomorphic and non-human infective. The origin of this cell line is detailed in Professor George Cross's web page (http://tryps.rockefeller.edu/).

The 427 derived 'S16' is a blood stream form cell line used for RNAi and has been engineered to constitutively express T7 polymerase and the tetracycline repressor protein from the non-transcribed spacer of the ribosomal DNA locus (Wirtz *et al.* (1999)). This allows the target mRNA transcript to be ablated upon induction with tetracycline. T7 polymerase expression is relatively deleterious to the cell, so this cell line was maintained in 2.5 μg ml^{-1} G418 to ensure retention of the T7 polymerase gene.

The 427-449 cell line is the Lister 427 transfected with the pHD449 construct (Wirtz *et al*. (1999)) used for inducible, ectopic expression in both blood forms and procyclic forms. It contains the tetracycline repressor gene integrated into the tubulin locus and transcription is driven by read-through transcription. This cell line is phleomycin resistant, stable, and does not need to be maintained in selective media.

The monomorphic DiD1 cell line was derived from *T.b.rhodesiense* GUP 2965.3 through successive rounds of selection against cells that are able to initiate differentiation (Tasker *et al.* (2000)).

AnTat 1.1 (van Meirvenne *et al.* (1975)) and the derived '90:13' cell line containing T7 polymerase and the tetracycline repressor protein (Engstler *et al*. (2004)) (a kind gift of Markus Engstler and Michael Boshart) are both *Trypanosoma brucei brucei*, pleomorphic, and are thought to be competent to complete the lifecycle through tsetse flies.

2.1.2 Culturing

Laboratory adapted blood form trypanosomes were cultured in HMI9 media in 5% CO_2 (Hirumi and Hirumi (1989)) using vented flasks (Greiner). Blood forms were maintained in logarithmic growth between 1×10^5 and 2×10^6 cells ml^{-1}. Cell counts were performed on a Beckman Coulter Z2 Coulter Particle Counter and Size Analyzer calibrated to 2.5 – 7.0 μm.

Pleomorphic blood forms were cultured in completed HMI9 media (Hirumi and Hirumi, (1989)) supplemented with 0.55% or 1.1% methylcellulose (for AnTat 1.1 or 90:13 cells respectively) and cell density was kept below 5×10^5 cells ml^{-1} for maintenance of logarithmic growth (McCulloch *et al*. (2004)). Cell counts were performed using a haemocytometer (see above).

Procyclic form trypanosomes were cultured in SDM79 (Cross and Manning (1973)) (Brun and Schonenberger (1979)) in sealed flasks. They were maintained in logarithmic growth between 1×10^6 and 2×10^7 cells ml^{-1}. Where procyclic forms were obtained from differentiated blood forms, SDM79 was supplemented with 10mM glycerol to maintain the cells in early procyclic form and to help form an established culture.

Cryopreserved stocks were made by centrifuging 10 ml of culture in logarithmic growth phase at 600g for 10 minutes. The pellet containing the cells was resuspended in 0.5 ml of supernatant, to which 0.5 ml of HMI9 supplemented with 14% glycerol was added, followed by gentle vortexing. This was placed into a cryovial at -80° C in a polystyrene insulator to allow slow freezing. After 2 days, the

cryovial was placed into a normal freezer box and stored at -80° C. Long term storage was in liquid nitrogen.

2.1.3 Purification of trypanosome cells from blood

Trypanosome cells were purified from mouse blood by passing them over a DEAE cellulose anion exchange column (Lanham and Godfrey (1970)). Briefly, DEAE cellulose was mixed with PSG pH 7.8 and allowed to settle in a glass column. 1 – 2 ml of blood was obtained by heart puncture and pippetted on to the top of the cellulose column. PSG was pippetted on to the column such that the trypanosome cells flowed through the column leaving the mouse blood cells within the matrix. The flowthrough was observed until it was no longer cloudy and subsequently centrifuged at 800 g for 10 minutes prior to use. Note that all mouse handling was performed by Deborah Hall or Keith Matthews, who are licensed to do this work.

2.1.4 Differentiation

2.1.4.1 *Cis*-aconitate induced differentiation

Unless otherwise stated, trypanosomes were differentiated from the blood form to the procyclic form by incubation in HMI9 or SDM79 media supplemented with 6 mM *cis*-aconitate, at 27°C in non-vented flasks.

Differentiation was assayed at various time-points post induction. Cells were either assayed using flow cytometry analysis or immunofluorescence to quantify the proportion of cells that were positive for the procyclic form marker, EP procyclin. Usually, 2×10^4 cells were analysed using flow cytometry, and 250 cells were analysed using immunofluorescence.

2.1.4.2 BZ3 induced differentiation

A stock solution of 67 mM BZ3 (chemical name: 3-(3,5-Dibromo-4-hydroxy-benzoyl)-2-ethyl-benzofuran-6-sulfonicacid-(4-(thiazol-2-ylsulfamyl)-phenyl)-amide) (Calbiochem) in dimethyl sulfoxide (DMSO) was added to trypanosomes in HMI9 or SDM79 at a final concentration of 150 μM (Szöor *et al*. (2006)). Cells

were incubated at 27°C and procyclin expression was monitored as described previously.

2.1.4.3 Pronase induced differentiation

Cells were washed in PSG pH 7.4 and resuspended in PSG supplemented with 4 U ml^{-1} Pronase (Sigma) for 10 minutes at 25°C. Cells were washed twice with PSG, resuspended in SDM79, and incubated at 27°C (Hunt *et al*. (1994)) (Sbicego *et al.* (1999)).

2.1.5 Transfection

Two methods based on electroporation were used when transfecting trypanosomes.

2.1.5.1 BTX

For each transfection, 10 ml of cells in mid-log growth were centrifuged at 600 g for 10 minutes. The cells were washed in half volume of ZFPMG (blood form parasites) or ZFPM (procyclic form parasites) at room temperature and resuspended at $1 – 2 \times 10^7$ cells ml^{-1}. DNA was prepared by restriction enzyme digestion of 15 μg of ultrapure plasmid overnight and purifying this through a nucleospin column (Macherey-Nagel). Electroporation was performed in 4 mm gap Electroporation cuvettes (Molecular BioProducts) using a BTX Electro Square Porator ECM830 set to 3 pulses of 1700V, 100 μs duration, at 200 ms intervals. Cells were recovered overnight in 10 ml HMI9 and plated out the next day in selective media at approximately 1×10^5 cells ml^{-1}. Transfectants required passaging 7 – 10 days post selection; if less than 30% of wells survived these were likely to be clonal.

2.1.5.2 Amaxa nucleofection

The Amaxa nucleofection has been reported to give 1000x higher yield of stable transfectants (Burkard *et al*. (2007)), and has been used with much success in this study.

2.1.5.2.1 Transfection of monomorphic cells

4×10^7 cells per transfection were centrifuged at 600g for 10 minutes and resuspended in 100 µl of Amaxa T cell buffer. DNA was prepared by overnight restriction enzyme digestion of 15 µg of ultrapure plasmid, purification through a Nucleospin column (Macherey-Nagel) followed by ethanol precipitation and resuspension in 5 µl of TE buffer. Electroporation was performed on the DNA mixed with the cells in the cuvettes supplied using the Amaxa nucleofector machine and programme X001. Cells were recovered overnight in 25 ml HMI9 and cultured thereafter in selective media. After 7 – 10 days the mock transfected cells were dead; if clonal cell lines were required it was often necessary to clone through limiting dilution in 24 well plates.

2.1.5.2.2 Transfection of pleomorphic cells

Transfection of pleomorphic cells is much more technically challenging than with monomorphic cells. The protocol was performed as described (McCulloch *et al.* (2004)). Cells from an early stage infection in mice (3 – 4 days, depending on the infection) with few, or no, intermediate or stumpy forms were purified from the mouse blood by centrifugation at 7000 rpm in a tabletop centrifuge, and using a yellow tip with the end cut off to take the white buffy coat. The red blood cell/ trypanosome mixture was transfected as described for monomorphic cells using the AMAXA nucleofection protocol (see 2.1.5.2.1). Cells were allowed to equilibrate in HMI9 media for 2 hours prior to being centrifuged and transferred at a concentration of approximately 5×10^6 cell ml^{-1} into HMI9 media supplement with 1.1% methyl cellulose. Half of the cells were plated into the top row of a 24 well plate (1 ml per well) and serial diluted 1:10 in methyl cellulose containing media, such that the final dilution of cells was 1:1000. The rest of the cells were allowed to remain in a vented flask. All cells were recovered overnight at 37°C in 5% CO_2 to allow expression of the drug resistance protein, and an equal volume of 2 × selective media (i.e. at double the normal concentration of drug) was added to the flask, and each well, such that the final concentration of selective drug was that normally used for selection of blood form parasites. Parasites in the flask were maintained at a concentration of < 5

x 10^5 cells ml^{-1} throughout selection. Non-transfected control parasites were monitored for cell death; 5 – 7 days post selection, transfectant wells, or parasites from the flask, were centrifuged and inoculated into a cyclophosphamide treated mouse. Blood from the resulting parasitaemia was harvested by heart puncture and freezer stocks were made as described (see Section 2.1.2) using a 1:1 ratio of blood: freezer mix.

2.1.5.2.3 Transient transfection of pleomorphic cells

Cells were taken from an early stage parasitaemia and purified by DEAE cellulose anion exchange (see Section 2.1.3). The cells were then centrifuged and transfected using 100 μg of uncut, ultra-pure plasmid DNA using the Amaxa nucleofection protocol (see Section 2.1.5.2), and allowed to recover in HMI9 media that had not been supplemented with methyl cellulose. Experiments were performed either 24, or 48, hours post transfection.

Note that the stage of parasitaemia was critical for experiments performed in this study. To induce RNAi prior to expression of stumpy specific proteins, a 3 – 4 day parasitaemia with < 2% intermediate or stumpy forms was required. Cells isolated too early in the infection did not differentiate to the stumpy form, and consequently died such that analysis could not be performed.

2.1.5.2.4 Drug concentrations

The blood form selective drug concentrations used in this study were: Hygromycin 2 μg ml^{-1}, phleomycin 2.5 μg ml^{-1}, G418 1 μg ml^{-1}, puromycin 0.5 μg ml^{-1} (Sigma).

2.1.6 Inducing cold shock

Stumpy form trypanosomes were purified from mice over a diethylaminoethyl cellulose (DE52) (Whatman) column, centrifuged, resuspended in HMI9, and incubated at 37°C, 5% CO_2 in vented flasks for 2 hours to equilibrate and recover. All solutions were pre-warmed to prevent inadvertent cold shock. Cells to be cold shocked were placed in non-vented culture flasks at 20°C for 18 – 24 hours.

2.2 Bacteria

2.2.1 Strains

Escherichia coli XL1- blue cells of genotype: *Sup*E44, hsdR17, *rec*A1, *end*A1, *gyr*A46, *thi*, *rel*A1, *lac,* F' [*pro*AB^{+}, *lac*I^{q}, *lac*ZDM15, Tn10(*tet*^{r})]. This strain gives good yields of high quality DNA and is suitable for beta-galactosidase based screening.

2.2.2 Transformation

100 µl of prepared cells were incubated with transforming DNA such that the volume of DNA did not exceed 5% of the final volume (i.e. 5 µl). After 30 minutes incubation on ice, the cells were subjected to heat shock of 42°C for 30 seconds, and were immediately placed on ice for 2 minutes. The cells were recovered in 300 µl LB or SOC for 1 hour at 37°C shaking at 220 RPM. The cells were then plated onto LB agar supplemented with 100 µg ml^{-1} ampicillin and incubated overnight at 37°C. Transformation efficiencies were approximately 10^{7} colony forming units (cfu) µg^{-1} of DNA.

2.3 Molecular Biology techniques

2.3.1 Polymerase Chain Reaction

The following represents a standard PCR reaction: 20% 5× buffer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 1 µM forward primer, 1 µM reverse primer, 0.0375 units µl^{-1}. A standard cyclical reaction for a non-proof reading polymerase such as GoTaq or Taq (Promega) was: 95°C denaturation for 30 seconds, 55°C annealing for 30 seconds, 72°C elongation for 1 minute kb^{-1} for 30 cycles. An initial denaturation of 95°C for 5 minutes and a final elongation of 72°C for 2 minutes kb^{-1} were also performed in addition to the standard 30 cycles. If a blend of a proof reading polymerase and non-proof reading was used, such as 'Expand Polymerase' (Roche), then the elongation times were doubled. Reaction buffers were purchased from Promega (proprietary composition) and oligonucleotide primers were purchased from Sigma Genosys.

2.3.2 Preparation and purification of DNA

2.3.2.1 DNA Preparation

Trypanosome genomic DNA was prepared by centrifuging a large number of live cells (> 5×10^7) at 600g for 10 minutes and lysing them in 300 µl TELT (see Appendix A) by gently pipetting up and down. DNA was purified from protein contaminants by adding 300 µl phenol: chloroform and gently mixing by inverting the tube, followed by centrifugation at 13K g for 10 minutes in a microfuge, and removal of the aqueous phase using a large bore pipette tip. This step was repeated to remove the last traces of protein. Salts from TELT were removed by addition of 900 µl ethanol and incubation at room temperature for 5 minutes; precipitated DNA was centrifuged at maximum speed in a microfuge and washed with 70% ethanol prior to resuspension in 100 µl of TE buffer.

Small-scale preparations of plasmid DNA were prepared using solutions I, II and III (Birnboim and Doly (1979)): 1 ml of overnight bacterial culture was centrifuged at 7K g for 5 minutes in a microcentrifuge. The supernatant was aspirated and the pellet resuspended in 100 µl solution I. 200 µl solution II was added to the cellular suspension and incubated at room temperature for 5 minutes to allow lysis. 150 µl of solution III was added to rapidly neutralise the reaction; the cells were then placed on ice for 10 minutes to allow precipitation of the proteins and chromosomal DNA. The cells were centrifuged at maximum speed for 10 minutes in a microfuge and the supernatant was re-centrifuged to remove the last of the cellular debris. Then, 900 µl of ethanol was added to the supernatant and incubated at -80°C for 1 hour to ensure plasmid precipitation. The DNA was recovered by centrifuging at 14K g for 10 minutes at 4°C, the supernatant aspirated and the DNA pellet washed with 500 µl of 70% ethanol. After air-drying for 5 minutes at room temperature, the pellet was resuspended in 50 µl of either ultra pure H_2O or TE, supplemented with 20 µg ml^{-1} RNase.

The QIAGEN Plasmid Midi kit was used for the larger preparations of pure plasmid DNA that were necessary for transfection. The DNA was prepared according to manufacturers instructions (QIAGEN Plasmid Purification Handbook, Third edition).

2.3.2.2 DNA purification

Salts and proteins can interfere with enzymatic reactions and transfections so it often necessary to purify DNA from contaminants. DNA was purified from agarose or from solution using a NucleoSpin Extract II kit (Machery-Nagel) as *per* the manufacturers instructions.

2.3.3 DNA restriction enzyme digestion

The following represents a typical restriction enzyme digest of DNA: 10% 10× buffer (NEB 1, 2, 3 or 4), 10% 10× BSA, 5% restriction enzyme(s), DNA. Reactions were incubated for < 2 hours at 37°C (with the exception of BclI with has an optimum digestion temperature of 50°C). Enzymes were purchased from either Roche, Qiagen or New England Biolabs.

2.3.4 DNA ligation

The following represents a typical ligation: 10% 10× ligase buffer (300mM Tris-HCl (pH 7.8), 100mM $MgCl_2$, 100mM DTT and 10mM ATP), 0.15 U μl^{-1} T4 DNA ligase, 5 ng μl^{-1} vector DNA, and the maximum amount of insert DNA in the remaining volume, to comprise a total of 10 µl. A 'no-insert' negative control was always performed, and reaction were incubated over-night at 4°C to allow ligation to occur.

2.3.5 DNA sequencing

Sequencing was performed by the Edinburgh School of Biological Sciences Sequencing Service (SBSS). Sequencing data was analysed using 4Peaks free software downloaded from http://mekentosj.com/4peaks/, in combination with an internet based alignment programme Multalin (http://bioinfo.genopole-toulouse.prd.fr/multalin/).

2.4 Northern analysis

All apparatus were cleaned by washing with detergent, followed by rinsing with ultra pure H_2O and 70% ethanol before air-drying.

2.4.1 Riboprobe preparation

A PCR reaction was performed using an insert cloned into pGEM T easy so that the amplicon contained T7 and SP6 polymerase binding sites. This amplicon was used as template in a digitoxin (DIG) RNA labelling reaction using the Roche RNA labelling kit following the manufacturer's instructions, so that a single stranded, DIG labelled, RNA probe was made that was anti-sense to the target mRNA transcript. Riboprobes were stored at -80°C.

2.4.2 Resolving trypanosome RNA on a denaturing agarose gel

An RNA denaturing gel was prepared by melting 1.2 g of agarose in 87 ml ultra pure H_2O followed by addition of 10 ml of 10× MOPS and 3 ml 37 % formaldehyde. The gel was allowed to set in a fume hood so that it had 16 wells for loading of RNA samples.

RNA was prepared from trypanosome cells using an RNeasy mini kit (QIAGEN catalogue #74106) according to the manufacturers instructions for animal cell RNA purification.

Approximately 4 µg of RNA in a maximum volume of 6 µl was mixed with 9µl formamide, 3µl formaldehyde (37%), 2µl 10xMOPS and 2 µl RNA loading buffer, denatured at 65°C for 5 minutes and loaded onto the gel immediately. The RNA was resolved at 150 V for 90 minutes in a 1× MOPS, stained with ethidium bromide and destained with water twice for 20 minutes. The RNA was visualised using a UV transilluminator and the ribosomal subunits photographed for use as an RNA loading control. In addition, the distance that the ribosomal subunits had migrated was measured using a ruler in order to calculate the size of any transcript that was detected.

2.4.3 Blotting trypanosome RNA

The RNA was blotted using a capillary blot system, where the bottom 2 cm of a plastic biohazard tray was filled with 20× SSC, and a gel tray was placed upside down on top of a small lunch box within the biohazard tray. A wick, made from 2× MME filter paper (Whatman), was wetted in 10x SSC and laid across the gel tray such that it made contact with the 20× SSC reservoir. The RNA gel was placed upside down on top of the wick, followed by a piece of positively charged membrane pre-wetted in 2× SSC cut to the size and shape of the RNA gel. 4 pieces of 2× MME paper were placed on top of the nylon membrane and the entire apparatus was wrapped in parafilm to prevent evaporation of the 20× SSC reservoir. A hole was cut in the parafilm to expose the gel stack, and 6 cm of absorbent tissue was placed on top of the gel stack with a 500 gram weight placed on top. The next day the nylon was removed to dry, and the RNA was then cross-linked to the membrane using a Uvitec cross linker at 0.12 joules for 120 seconds.

2.4.4 Probe hybridization and detection

1 μl of the riboprobe was denatured by boiling for 5 minutes in hybridisation buffer and then used to probe the nylon membrane overnight at 68°C in 8 ml of hybridisation buffer. Using freshly cleaned boxes for each step, the membrane was washed twice for 30 minutes at hybridisation temperature in 2xSSC/0.1% SDS, and once for 30 minutes in 0.5xSSC/0.1% SDS. The membrane was then blocked for 1 hour at room temperature in Maleic acid with 1% block prior to incubating with 1 μl anti-DIG antibody in 50 ml of Maleic acid containing 1% block, for 30 minutes. The membrane was washed 3 x 10 minutes in Maleic Acid Buffer with 0.3% Tween-20 before soaking for 2 minutes in detection buffer (100mMTris-HCl, 100mM NaCl pH 9.5).

The membrane was incubated with 1 ml 1× CDP star (New England Biolabs) in bag W for 2 minutes before removing the CDP star and incubating the membrane in sealed hybridisation bag W for 15 minutes. The membrane was exposed to X ray film inside a cassette for 5 minutes to determine the required exposure time before development.

2.5 Southern Analysis

2.5.1 DNA probe synthesis

A DIG DNA labelling kit (Roche) was used to synthesise a DNA probe according to the manufacturer's instructions. The template for this reaction was a PCR product with homology to the target sequence.

2.5.2 Resolving trypanosome genomic DNA on an agarose gel

5 μg of trypanosome genomic DNA was digested overnight with a single restriction enzyme and resolved on a 1% agarose gel in 1× TAE buffer at 150 V for 90 minutes. The gel was stained with ethidium bromide, destained with water, and visualised using a UV transilluminator.

2.5.3 Genomic DNA blotting

The gel was de-purinated with 0.25 M HCl^- for 30 minutes, denatured with Denaturing Solution (1.5 M NaCl, 0.5 M NaOH) for 30 minutes, and neutralised with Neutralising Solution (1.5 M NaCl, 1 M Tris, pH 7.4) for 30 minutes. The DNA was then blotted using a capillary blot apparatus and cross linked to the nylon membrane as for a Northern analysis (described above).

2.5.4 Probe hybridization and detection

The membrane was blocked in 10 ml of hybridisation buffer (5% dextran sulphate, 5×SSC, 10% liquid block solution (Pharmacia), 0.1% SDS) for 1 hour at 65°C in a rotating hybridisation tube. The hybridisation buffer was replaced with fresh, pre-warmed buffer, and 250 ng of DNA probe denatured by boiling at 100°C for 5 minutes was added to the membrane hybridisation buffer and incubated overnight.

The membrane was washed twice with 50 – 100 ml of pre-warmed 1×SSC, 0.1%SDS for 30 minutes at 65°C, and then twice with 50 – 100 ml of pre-warmed 0.1×SSC, 0.1% SDS at 65°C. The membrane was then blocked in 50 ml blocking buffer at room temperature for 1 hour, and probed with 2 μl α-DIG antibody diluted in 50 ml buffer A. Unbound conjugate was removed by washing 3 times for 15 minutes with 0.3% Tween in buffer A (100 mM Tris, 300 mM NaCl, pH9.5). Signal was detected

by incubating the membrane with 1 ml 1× CDP star (New England Biolabs) for 2 minutes, removing excess CDP star in sealed hybridisation bag W and exposing X ray film to the membrane in a photographic cassette.

2.6 Western analysis

2.6.1 Preparation of protein sample

A known quantity of live trypanosomes in culture was pelleted at 600 g for 10 minutes, washed in PBS and resuspended in protein sample buffer at a concentration of 3×10^6 trypanosomes in 5 µl. Cells were lysed and genomic DNA was sheared by either passing the cells through a 0.5 mM gauge needle (BD Microlance) or, where the sample volume was very little, by placing in a room temperature water bath sonicator (Bandelin Sonrex RK100 H) for 20 seconds. Cell lysate was re-suspended in Laemmli protein sample buffer (62.5 mM Tris-HCl pH6.8, 2% SDS, 10% glycerol, 0.002% Bromophenol blue, 5% β-mercaptoethanol) at 3×10^6 cells *per* 10 µl. Protein sample was kept on ice wherever possible, and were not boiled unless otherwise stated in order to prevent hydrophobic proteins from precipitating.

2.6.2 Preparation of SDS gel

The following comprises the components of a 12 % acrylamide SDS gel: 12 % acrylamide-bisacrylimide mix, 375 mM Tris pH8.8, 0.005 % ammonium persulphate (APS) (w/v), 0.005 % tetramethylethylenediamine (TEMED), where the APS and the TEMED were added immediately prior to pouring the gel. The gel was poured in a Bio-Rad PROTEAN II casting apparatus, and allowed to set for approximately 15 minutes to give a 0.75 mm thick gel. A stacking gel (375 mM Tris pH6.8, 6.8% acrylamide, 0.1% SDS, 0.056% ammonium persulphate, 0.0056% TEMED) was poured on top of the resolving gel and allowed to set with a 10 or 15 well comb inserted such that samples of 10 - 20 µl could be loaded.

2.6.3 Resolving proteins on an SDS polyacrylamide gel

Samples were loaded such that protein from 3×10^6 cells was resolved in each lane. The gel was run at 100 V for approximately 2 hours on ice in 1 × running buffer (25

mM Tris-HCl pH8.3, 192 mM glycine, 0.1% SDS). If the protein was to be visualised on the gel then the gel was soaked in Coomassie blue stain (50% methanol, 10% acetic acid, 0.1% Coomassie Brilliant blue) for 10 minutes, and destained several times with Destain solution (40% methanol, 10% acetic acid) until it was possible to distinguish the different protein bands.

2.6.4 Blotting proteins

2.6.4.1 Semi-dry transfer

Proteins were blotted using either a semidry blotter (BIO-RAD Trans-blot SD semi-dry transfer cell) and power-pack (Bio-Rad Power-PAC 200) onto nitrocellulose membrane (Whatman Schleicher and Schuell, PROTRAN, pore size 0.45 μm) according to manufacturer's instructions. Briefly, the acrylamide gel, nitrocellulose cut to the correct size and extra thick filter paper were equilibrated in cold transfer buffer (2.5 mM Tris, 15 mM Glycine, 0.02% SDS, 20% methanol) for 5 minutes and placed sequentially onto the semidry blotter in the following order: 3 mm extra thick filter paper (Whatman), nitrocellulose, acrylamide gel, extra thick filter paper. The blot was carefully rolled between as each layer was added to ensure no air bubbles interfered with the transfer, and then transferred at 15 V for 60 minutes. The proteins were visualised by staining with Ponceau stain (0.4% Ponceau S, 3% trichloroacetic acid (TCA)) for 10 minutes, and de-staining several times with ultra pure water.

2.6.4.2 Wet transfer

Protein were blotted using a wet blotting system (Bio-Rad) according to manufacturers instruction's. Briefly, the gel, PVDF membrane (Immobilon-P, Millipore), and 3MM filter paper (Whatman) were equilibrated in wet transfer buffer (25 mM Tris-HCl, 192 mM glycine, 20% methanol) for at least 30 minutes. Each layer was placed onto the wet blotting apparatus in the following order, with careful rolling in between each layer to ensure there were no bubbles: 2 layers of 3MM filter paper, the acrylamide gel, the PVDF membrane, and the 2 layers of filter paper. The 'sandwich' was placed into the transfer apparatus, which was filled with chilled wet transfer buffer, and left to transfer at 30V overnight at 4°C.

2.6.4.3 Blocking and staining blot

See Appendix A for antibody concentrations.

Blots were blocked overnight at 4 °C with 5 % non-fat powdered milk (Marvel Original dried skimmed milk) in 1× PBS or LI-COR Odyssey block diluted 1:1 in 1× PBS, and then stained with the primary antibody in either 5 % milk: PBS or LI-COR Odyssey block diluted 1:1 in 1× PBS(0.05% Tween) by gently rocking for 1 hour at room temperature or overnight at 4°C. The blot was then washed 5 × 3 minutes by rocking gently with excess 0.05 % TWEEN in 1× PBS, stained with the appropriate secondary antibody in 5% milk: PBS or LI-COR Odyssey block diluted 1:1 in 1× PBS(0.05% TWEEN) for 1 hour at room temperature. The washes were then repeated.

2.6.4.4 Visualising the chemiluminescent signal

The blot was incubated with either Enhanced chemiluminescence (ECL) substrate (GE Healthcare, Amersham) for 1 minute or high sensitivity ECL (Millipore Immobilon Western Chemiluminescent HRP Substrate) for 5 minutes depending on the sensitivity that was required. Resulting bands were visualised by exposing X ray film to the blot in a photographic cassette.

2.6.5 Quantitative Western blots

The use of the LI-COR Odyssey system allowed quantitation of the signal obtained from Western blots. This system makes use of a secondary antibody conjugated to a fluorochrome, which is subsequently detected using an infrared laser. Co-staining with a tubulin loading control using secondary antibodies conjugated to different fluorochromes with non-overlapping emission spectra allowed relative protein abundance to be quantified.

The principle difference in utilising this system is the use of secondary antibodies. Red anti-rabbit (product #: 926-32221) and green anti-mouse (product #:926-32210) were purchased from LI-COR and used in place of HRP-conjugated secondary

antibodies. Blots were scanned using the LI-COR Odyssey Imager, and the signal was analysed in Microsoft Excel relative to the tubulin signal.

2.6.6 Western blot of PAD proteins

Hydrophobic proteins are notoriously difficult to detect due to their tendency to precipitate, and therefore not enter the acrylamide get. The following represents the fully optimised protocol for PAD protein detection. The principle difference from the standard Western blot protocol was that the protein sample was maintained at a low temperature at all times.

A known concentration of cells was centrifuged, resuspended in 1 ml of PBS, and re-centrifuged. The supernatant was aspirated, and resuspended at 3 x 10^6 cells *per* 10 µl of protein sample buffer containing β-mercaptoethanol by 'grating' the eppendorf along an eppendorf rack. Genomic DNA was sheared by sonication in a sonicator water bath, and the sample was then placed upon ice. 10 µl of protein sample was resolved on a 10% polyacrylamide gel at 100V for approximately 4 hours using chilled buffers in a 4°C fridge. Gels were blotted onto PVDF Immobilon-P (Millipore) using a wet blotting system (Bio-RAD) with chilled buffers in the fridge as per the manufacturer's instructions. Membranes were subsequently treated as for a standard quantitative, or chemiluminescent, Western blot (see Section 2.6.4.3).

2.7 Immunofluorescence

Immunofluorescence was used for two different purposes in this study. The first was to stain for a life cycle development marker such as EP Procyclin in order to quantify a population's progression through differentiation. The second was to localise a protein within the cell. As the objective was different in each case, the cells required different fixation conditions.

Stained slides were visualised using either a 'Zeiss Axioskop 2 plus' microscope, or a Leica SP6 confocal microscope. Images were analysed in either Volocity (Improvision) (for the confocal data), Adobe Photoshop or ImageJ.

2.7.1 Methanol fixation

Methanol fixation preserves the cellular morphology of the fixed cell, but disrupts the internal cellular structure. Hence this methanol fixation is not optimal for protein localisation.

Approximately 10^6 cells from culture were centrifuged at 600 g for 10 minutes, resuspended in approximately 30 µl of supernatant, and spread across a glass slide (Menzel-Glaser Microcope slides 76 × 26 mm). After approximately 30 minutes of air-drying the cells were fixed by immersing in cold methanol for at least 10 minutes, and stored at -20 °C.

All stains and blocks were performed in a humidity chamber to prevent evaporation. Cells were re-hydrated in excess PBS and blocked with 100µl 20 % foetal bovine serum (Invitrogen) in PBS for 30 minutes, stained with 100µl primary antibody diluted in 2 % BSA PBS for 45 minutes and washed 3 × 5 minutes in excess PBS, re-stained with 100µl secondary antibody diluted in 2 % BSA PBS for 45 minutes and the washes repeated. Finally, the cellular DNA was stained with 10 µg ml^{-1} 4',6-diamidino-2-phenylindole (DAPI) and the slides washed once with excess PBS and mounted in 10 % PDA in MOWIOL using a coverslip (Scientific Laboratory Supplies). Cells were kept at 4°C to allow the MOWIOL to set and the PDA diffuse into the cell.

2.7.2 Paraformaldehyde fixation

The protocol was adapted from published protocols (Field *et al.* (2004)). Briefly, 2 × 10^6 cells were pelleted at 600 g for 10 minutes, washed in vPBS, and resuspended in 100 µl vPBS. 100 µl of 6 % paraformaldehyde was added to the cells for fixation. After exactly 10 minutes, 5 ml of vPBS was added to the cells, which were pelleted gently, and allowed to settle on a poly-lysine coated slide (VWR International) for 20 minutes before being washed in excess 1× PBS. Cells were permeabilised for 10 minutes using 0.05% TritonX, blocked in 20 % foetal calf serum in vPBS for 45 minutes, and stained with the primary antibody in 20 % foetal bovine serum in vPBS

for >1 hour. Subsequent to washing in excess PBS, the cells were then stained with the secondary antibody in 20 % FBS: vPBS for 1 hour, washed in excess PBS 3 × 5 minutes after which the cellular DNA was stained with a 1 μg ml^{-1} working stock of DAPI. The cells were washed again 3 × 5 minutes in excess PBS and mounted using 10 % PDA in MOWIOL. Cells were kept at 4°C to allow the MOWIOL to set and the PDA permeate into the cell.

2.8 Flow cytometry

Flow cytometry was used primarily as a way of assaying the progression of differentiation to the procyclic form of a population of cells.

2.8.1 Fixing cells

Between 2 – 5 × 10^6 live cells from culture were fixed by centrifuging at 600 g for 5 minutes, washing with PBS and re-suspending in 2 % formaldehyde/ 0.05 % glutaraldehyde fix for a minimum of 1 hour at 4 °C (Roditi *et al.* (1989)). Cells were kept up to 72 hours in fix prior to staining; note that best fixation occurred when cells were allowed to fix for at least 24 hours, as judged by the forward and side scatter.

2.8.2 Blocking and staining

Washes were performed by centrifuging cells at 600 g for 6 minutes and re-suspending in PBS and re-centrifuging as before. Stains were performed by incubating the cellular suspension in 200 μl of antibody diluted in 2 % BSA in PBS. Cells were washed twice to remove the last traces of fixative prior to staining with the primary antibody for 1 hour at 4°C, washed twice and stained with the secondary antibody for 1 hour at 4°C. Cells were washed once again before analysis.

2.8.3 Use of the FACS machine

Cells were analysed using FACSCalibur FACS machine (BD). In all cases, an unstained negative control was used to determine the level of auto-fluorescence. In addition, a 'secondary only' control was included to determine the level of non-specific binding of the fluorochrome-conjugated antibody. Wherever possible a

positive and negative control sample was also used to help calibrate the settings on the FACS machine so that positive and negative cells could be distinguished.

2.8.4 Analysis of data

Data was analysed using FlowJo (Tree Star).

2.9 Citrate transport assay

2.9.1 Synthesis of PAD1 RNA

PAD1 was cloned into the EcoRI and HindIII digested *Xenopus* expression vector pGHJ (Liman *et al*. (1992)) (Bröer *et al.* (1997)) such that a Kozak sequence was cloned into the 5' end of PAD1, and an HA epitope tag would be synthesised on the C-terminus of the translated protein. A small scale preparation of plasmid DNA was linearised, and synthetic PAD1 RNA was made using RNA polymerase in the presence of 1 mM rATP, 1 mM rCTP, 1 mM rUTP, 0.2 mM rGTP and 1 mM of the cap analogue m7G(5')ppp(5')G according to manufacturer's instructions (Promega, Madison, WI). RNA was purified after DNase treatment through a nucleotide purification column (QIAGEN), eluted in 2 × 50 µl of elution buffer, and ethanol precipitated for resuspension at approximately 1.5 mg ml^{-1}.

2.9.2 Preparation of *Xenopus* oocytes for micro-injection

Xenopus laevis frogs were operated on by Rosa Marchetti. Briefly, the frog was made unconscious by submersion in a 0.1% 3-aminobenzoic acid ethyl ester solution (Sigma, St. Louis, Mo, USA) and placed on ice for surgery. Oocytes were removed after surgery on the abdomen of the frog with a scalpel to remove a large number of oocytes (as many as 100). Oocytes were incubated with trypsin for an hour at 4 °C to remove extraneous tissue and to allow for easier injection.

2.9.3 Microinjection of *Xenopus laevis* oocyctes

Glass capillaries (WPI, Sarasota, FL, USA) with a long shank were pulled using a normal puller to make a micro-injection needle. A WPI microinjector (WPI, Sarasota, USA) was used to inject purified PAD1 cRNA into oocytes held in place on a 35 mm petri dish with a polypropylene mesh glued to the bottom to fix the

oocytes. Oocytes were incubated at 16°C in ND96 buffer (96 mM NaCl, 2 mM KCl, 1.8 mM CaCl2, 1 mM MgCl2, 5 mM HEPES, pH 7.4) for 5 – 7 days to allow recovery of the oocytes, and the translation and localisation of PAD1 to the cell surface membrane.

2.9.4 Citrate transport assay

10 oocytes were used per condition. 10 oocytes were transferred into a 5 ml tube and flushed twice with 4 ml ND96 at either pH7, 8, or 9. The supernatant was then removed completely. Thereafter, 100 µl of transport buffer containing citrate and radiolabelled ^{14}C citrate was added to each oocyte and incubated for either 30 or 60 minutes. 3 ml of ice cold transport buffer was added to stop transport, and the oocytes were washed 3 × with 4 ml ice cold transport buffer. 1 ml of ice cold transport buffer was added and the oocytes were removed into a scintillation vial and vortexed in 200 µl of 10 % SDS buffer to lyse the cell. Next, 1.5 ml of scintillation fluid was added and each vial was put into a liquid scintillation counter to detect the quantity of radiolabeled citrate taken up by the cell.

2.9.5 Analysis of data

Analysis was performed using excel to calculate the average radioactivity (counts per minute) by the oocytes.

2.10 Plasmid constructs used in this study

Trypanosome transgenic technology utilizes plasmid constructs that allow inducible ectopic expression, ablation and deletion of trypanosome genes. This study was heavily reliant on the constructs described.

2.10.1 pHD451

Described in Biebinger *et al.* (1997), this is an ectopic expression plasmid primarily designed for expression in procyclic form trypanosomes that integrates into the ribosomal spacer, and utilizes a procyclin promoter to express the *trans*-gene. Due to incompatible cloning sites, it was necessary to insert a XhoI restriction enzyme site between the HindIII and BamHI sites using annealed primers #24 and #25 (see

Appendix A, and Section 5.2.1.1). Low expression in bloodstream forms made this construct non-optimal for over-expression of PAD1 in monomorphic bloodstream forms.

In addition, a version of this construct that contains a TY epitope tag sequence downstream of the insertion site was used to allow synthesis of an epitope tagged protein and subsequent detection using BB2 antibody in transfected parasites.

2.10.2 pHD617 and pHD617 (PuroR)

Described in Biebinger *et al.* (1997), pHD617 is optimized for expression in bloodstream form trypanosomes. pHD617 confers hygromycin resistance and was therefore not compatible for conditional rescue of the hygromycin resistant knockout parasites. For this reason, a puromycin resistance version of pHD617 was made. Briefly, pHD617 was modified to contain an NdeI at the 5' end of the resistance gene by digesting with XhoI and NdeI and inserting an amplicon containing the EP procyclin promoter from pHD451 using primers #92 and #93. An amplicon containing the puromycin resistance cassette from pHD1034 was cloned into NdeI and AvrII sites of the modified pHD617 using primers #94 and #95 to give pHD617 (PuroR). See Appendix A for map and the list of primers.

2.10.3 pHD449

pHD449 (Biebinger *et al.* (1997)) contains the tetracycline repressor protein and, upon transfection, integrates into the tubulin gene array. Transcription is constitutive and reliant on read-through transcription through the tubulin gene locus.

2.10.4 pALC14

pALC14 (Bochud-Allemann and Schneider (2002)) was a kind gift from Andre Schneider (University of Bern, Switzerland). pALC14 contains a puromycin resistance cassette and allows inducible expression of a stem loop RNA transcribed from the ribosomal spacer region to give ablation of a target mRNA transcript (see Appendix A for a schematic of this construct). An amplicon was first ligated into the XhoI-BamHI cloning site, and then the HindIII-NdeI-XbaI multiple cloning site, such that the 2 amplicon sequences were separated by a 460 nucleotide 'stuffer'

sequence and in the reverse complement orientation (see Section 6.2.1 for details). When generating pALC14 constructs to target PAD1 and the entire PAD array, BamHI was not used as this site is already present within the amplicon. For this reason, BclI was used in its place because fragments generated using this enzyme can be cloned into BamHI restriction sites. Two pALC14 based constructs were used in this study: 'array:pALC14' using primers #106 and #109 that was designed to ablate all of the genes in the PAD array, and 'PAD1: pALC14' made using primers #108 and #109 that was designed to only ablate PAD1 mRNA.

2.10.5 Gene deletion constructs

Gene deletion was performed using drug resistance cassettes flanked by regions of homology to sequences adjacent to the region to be deleted. The constructs were generated from existing constructs containing Neomycin and Hygromycin resistance cassettes (Hendriks and Matthews (2005)) in order to generate cell lines that have PAD1, and the entire 28 kb array, deleted. The neomycin resistance construct contains a tetracycline repressor gene to allow inducible expression of a rescue gene.

2.10.5.1 PAD1 deletion constructs

The first round PAD1 deletion construct was made by amplifying regions in the 5' (primers #67 and #68) and 3' UTR (primers #63 and #64) of PAD1 from genomic DNA and inserting them into the cloning sites flanking the neomycin resistance gene and tetracycline repressor gene so that the final construct has (5' → 3') PAD1 5' UTR – tetracycline repressor gene –neomycin resistance gene - PAD1 3'UTR. The second round PAD1 deletion construct was made in the same way using homologous regions (primers #46 and #45, #65 and #66) nested within the first round knockout, so that the final construct has (5' → 3') PAD1 5' UTR – hygromycin – PAD1 3' UTR. Both constructs were digested with KpnI and SacI and purified using a nucleospin column prior to transfection. The list of primers can be found in Appendix A, and an explanatory diagram can be found in Section 6.4.1.

2.10.5.2 PAD array deletion constructs

In this case, the region to be removed was much larger so it was judged that much bigger regions of homology would be necessary, and consequently it was not possible to perform a 'nested' knock out. A 2.5 kb region of homology was cloned into the 5' insertion site of the neomycin resistance gene deletion construct (primers #75 and #45) and a 1.6 kb region of homology was cloned into the 3' insertion site. To make the second round deletion construct, the tetracycline repressor and neomycin resistance cassettes were removed by digestion with XhoI and NotI, and replaced with an amplicon containing the hygromycin resistance cassette from the second round deletion construct made using primers #87 and #88.

2.10.6 pGEM T easy

pGEM T easy is a commercially available plasmid (Promega) and used for PCR product cloning without digestion, exploiting A overhangs generated by Taq and Taq blended polymerases. This was used in riboprobe synthesis and DNA probe synthesis.

2.10.7 pGem-He-Juel

pGem-He-Juel, or pGHJ (Liman *et al*. (1992)) (Bröoer *et al.* (1997)), is a *Xenopus* expression vector commonly used for *Xenopus* uptake assays, and contains a Kozak sequence and β tubulin intergenic regions that ensure high expression.

Chapter 3

A Bioinformatics Analysis of the PAD Genes

3 A Bioinformatics analysis of the PAD array

The first stage in characterising a newly identified gene is conventionally a bioinformatics analysis. Bioinformatics is the application of computing technology to rationalise large amounts of biological data, allowing cryptic information to be revealed. A number of programmes and algorithms have been developed that analyse the primary sequence of DNA, RNA or a protein in order to build a bioinformatics profile of the gene of interest. The nature of these programmes varies depending on the type of analysis, but some examples of their use are:

- Inference of evolutionary relationships based upon DNA sequence divergence
- Elucidation of protein function based upon sequence similarity to previously characterised proteins
- Using the known physiochemical properties of amino acid residues to annotate a protein (e.g. a sequence of hydrophobic residues flanked by hydrophilic residues often constitutes a transmembrane domains)
- Identification of motifs and domains of known function (e.g. zinc finger motifs, kinase domains, phosphatase domains)
- Structural prediction algorithms to predict secondary and tertiary structures of RNA and proteins
- Identification of short peptide sequences known to mediate localisation to specific cellular compartments (e.g. nuclear localisation signals).

Some of these programmes were utilised to build a bioinformatics profile of the PAD genes and proteins.

3.1 The genomic context of the PAD genes

The *Trypanosoma brucei* haploid genome is made up of over 9000 predicted genes in 11 chromosomes arranged over approximately 26 Mb (Berriman *et al.* (2005)). These genes are arranged as directional gene clusters of up to 1 Mb that are probably transcribed as large polycistronic transcription units (Sather and Agabian (1985)). Despite early divergence of the 'tri-tryps' (Stevens and Gibson (1999)) (consisting of *Trypanosoma cruzi, Leishmania major* and *T. brucei*) there is a very high degree of synteny between the sequenced genomes (approximately 68% for *T. brucei* and 75% for *L. major*) (El-Sayed *et al.* (2005)), especially within the directional gene clusters. Genes that are unique to a trypanosomatid genome are thought to be mostly associated with specific parasite life cycle roles (such as entry into the macrophage in the case of *L. major*, or antigen proteins in the case of *T. brucei*) or to have distinct metabolic or physiological roles. Interestingly, 43% of genes that represent breaks in synteny are found in, or near, strand-switch regions between the directional gene clusters; this may represent either a positive selection pressure maintaining synteny, or may be associated with the mechanism of recombination between homologous chromosomes (El-Sayed *et al.* (2005)).

The PAD genes are arranged as an 8 gene cluster over 27 kb of chromosome 7 (referred to as the PAD array) (see Figure 3.1-1) approximately 600 Kb from the closest telomere. They are situated at the end of a directional gene cluster (DGC) approximately 300 Kb in length and there are no syntenic PAD orthologues found in *L. major* or *T. cruzi* (see Figure 3.1-2). A ClustalW alignment of the closest orthologs found by a TBLASTX search of the *L. major* and *T. cruzi* genomes using the PAD1 protein sequence shows 39% amino acid identity between the 3 translated genes (Figure 3.1-3). The *T. cruzi* homologue (Tc00.1047053506551.10) shows closest similarity with 56% amino acid identity to PAD1, and the *L. major* homologue (LmjF11.0650) has slightly less identity at 52% (see Appendix A for the individual alignments with PAD1). A reciprocal BLASTN and TBLASTX search, whereby the *T. cruzi* and *L. major* closest homologue nucleotide sequences were used to interrogate the *T. brucei* genome data, returns members of the PAD array as

the closest match (data not shown), thus verifying that PAD1 is the closest ortholog to these genes within the *T. brucei* genome.

Interestingly, the first and third closest orthologs in *L. major* (LmjF11.0650 and LmjF11.0660 respectively) are members of an array of 6 homologous genes on chromosome 11 (see Figure 3.1-6), all of which show closest homology to PAD genes in a reciprocal TBLASTX against the *T. brucei* genome (data not shown). The second closest ortholog (LmjF13.1690) is situated at the 5' end of an incompletely sequenced direction gene cluster, and does not appear to be part of a duplicated gene array (data not shown).

Analysis of the *T. cruzi* orthologs was complicated by the incomplete sequencing of this genome. The first and third best scoring TBLASTX hits (Tc00.1047053506551.10 and Tc00.1047053510069.20 respectively) have no genomic context or adjacent genes annotated in the GeneDB database. The other high scoring *T. cruzi* TBLASTX hits were found to be on DNA fragments that contained 2 or more adjacent PAD orthologues. For example, the second closest PAD1 orthologs (Tc00.1047053508799.270) was situated between the seventh and thirteenth closest orthologs (Tc00.1047053508799.260 and Tc00.1047053508799.280 respectively) (Figure 3.1-5 shows an alignment of these proteins with the PAD1 protein, and Figure 3.1-6 shows the genomic context of these genes), and the fourth closest ortholog (Tc00.1047053509713.30) was situated downstream of the twelfth and tenth closest orthologs (Tc00.1047053509713.10 and Tc00.1047053509713.20 respectively). See Appendix A for a list of the *T. cruzi* orthologs. It seems likely that the *T. cruzi* PAD orthologs identified form part a duplicated gene cluster analogous to the *T. brucei* PAD array. As with *L. major*, the genes upstream of the *T. cruzi* PAD orthologs show no homology with the genes upstream of the PAD array in *T. brucei*; hence the *T. cruzi* and *L. major* PAD array equivalents are non-syntenic with the *T. brucei* PAD array.

Thus, it is likely that expansion of an ancestral PAD gene occurred prior to divergence of the salivarian trypanosomes approximately 300 – 500 million years ago (Gibson and Stevens (1999)) and that the additional two genes in the *T. brucei*

genome represent a more recent gene duplication. Alternatively, expansion of all 8 genes in the array could have occurred before the tri-tryp divergence and *L. major* (and possibly *T. cruzi*) has lost 2 genes through genomic re-arrangement.

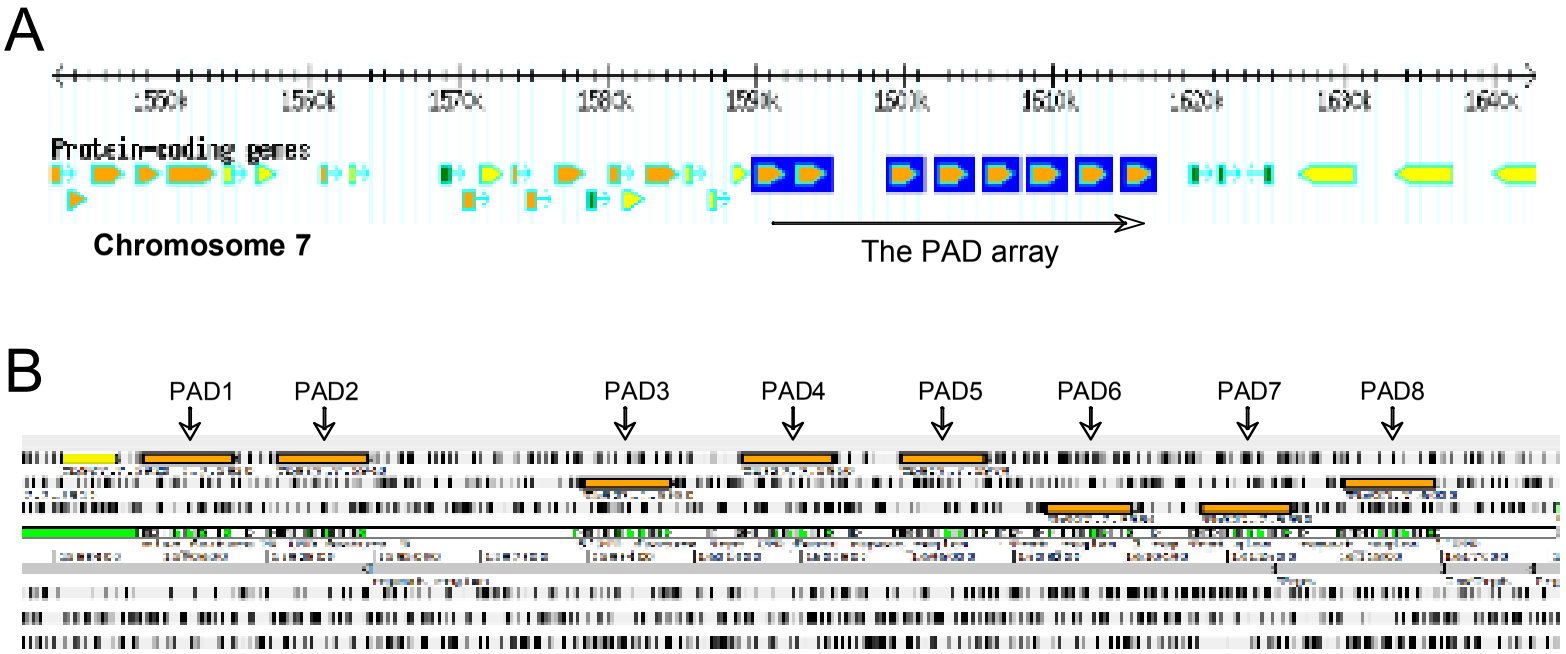


Figure 3.1-1 The genomic context of the PAD array (A) The genomic context of PAD1 was visualised using the genome browser integrated into the GeneDB database. The PAD array is situated at the end of a 300 Kb directional gene cluster on chromosome 7. The PAD genes are in blue highlight as indicated. Other genes are depicted in either orange or yellow. The direction of transcription is indicated by the arrow. (B) A higher resolution map of the PAD array was displayed using the ARTEMIS program (Rutherford *et al.* (2000)). There are 8 PAD genes in the same orientation covering 27 Kb of the genome with no intervening open reading frames. There is approximately even distribution of the genes, with the exception of the larger gap between PAD2 and 3.

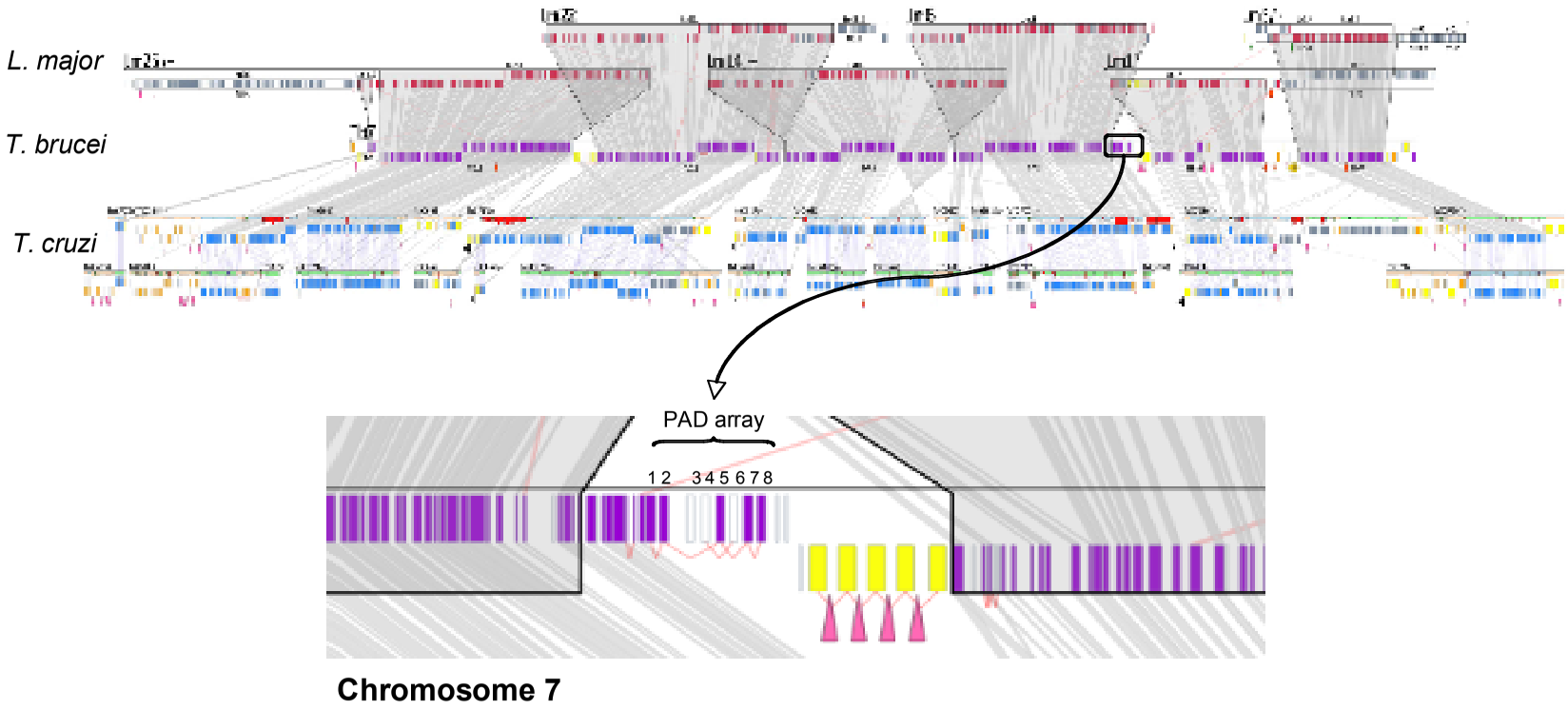


Figure 3.1-2 Synteny in the PAD region of the trypanosomatid genomes. Syntenic chromosomes of *T. cruzi* and *L. major* were aligned with *T. brucei* chromosome 7. The PAD array represents a break in synteny and is not found in either *T. cruzi* or *L. major*. The PAD array is not found at syntenic genomic positions in either the *Leishmania* or *Trypanosoma cruzi* genomes (Figure adapted from El-Sayed *et al.* (2005)). Regions of synteny are indicated by grey lines, genes are indicated by yellow, purple or clear boxes.

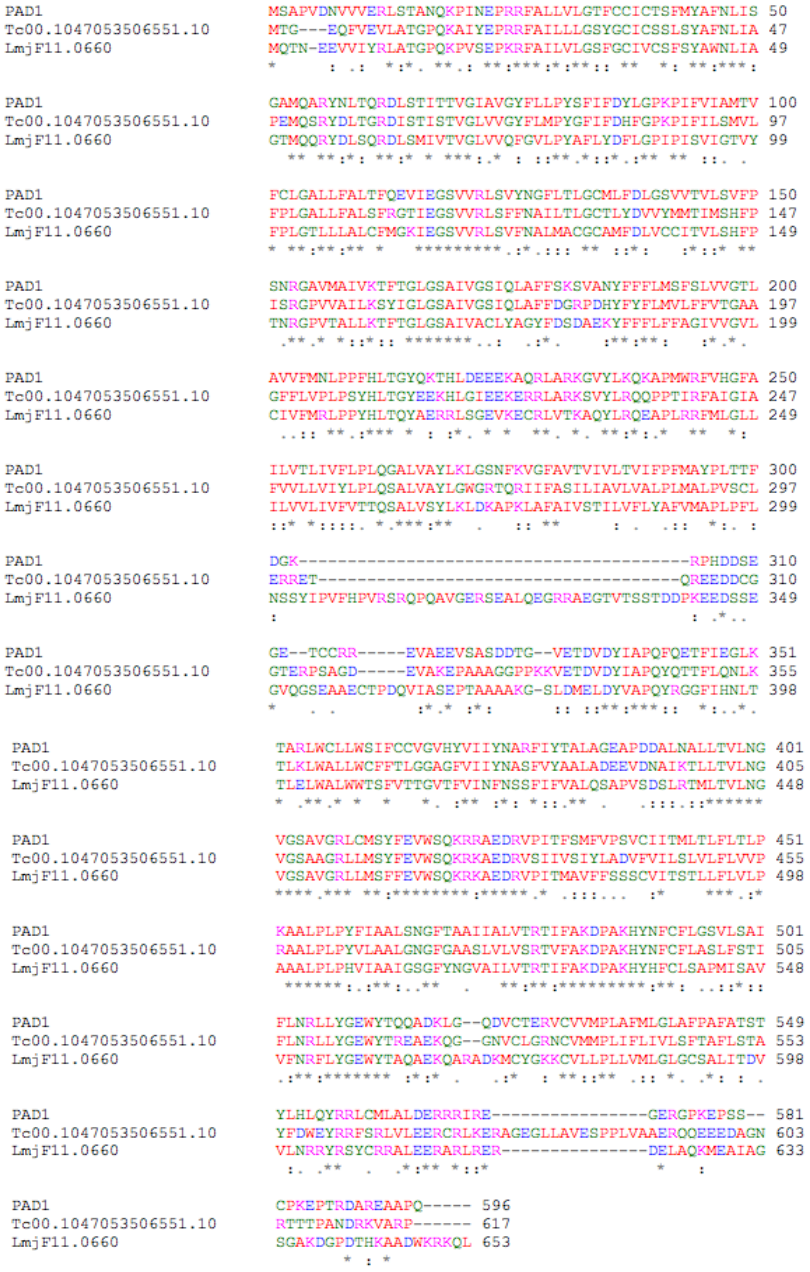


Figure 3.1-3 A protein alignment of PAD1 and the closest orthologs in *L. major* and *T. cruzi*. Orthologs were found using PAD1 as a query sequence in a TBLASTX search of the *L. major* and *T. cruzi* genomes. The orthologs were aligned using clustalW using the default settings. The alignment shows that the *T. cruzi* and *L. major* proteins have 39% amino acid identity with PAD1. The asterisks, colons and dots indicate perfect, very high and high conservation between the 5 proteins, respectively.

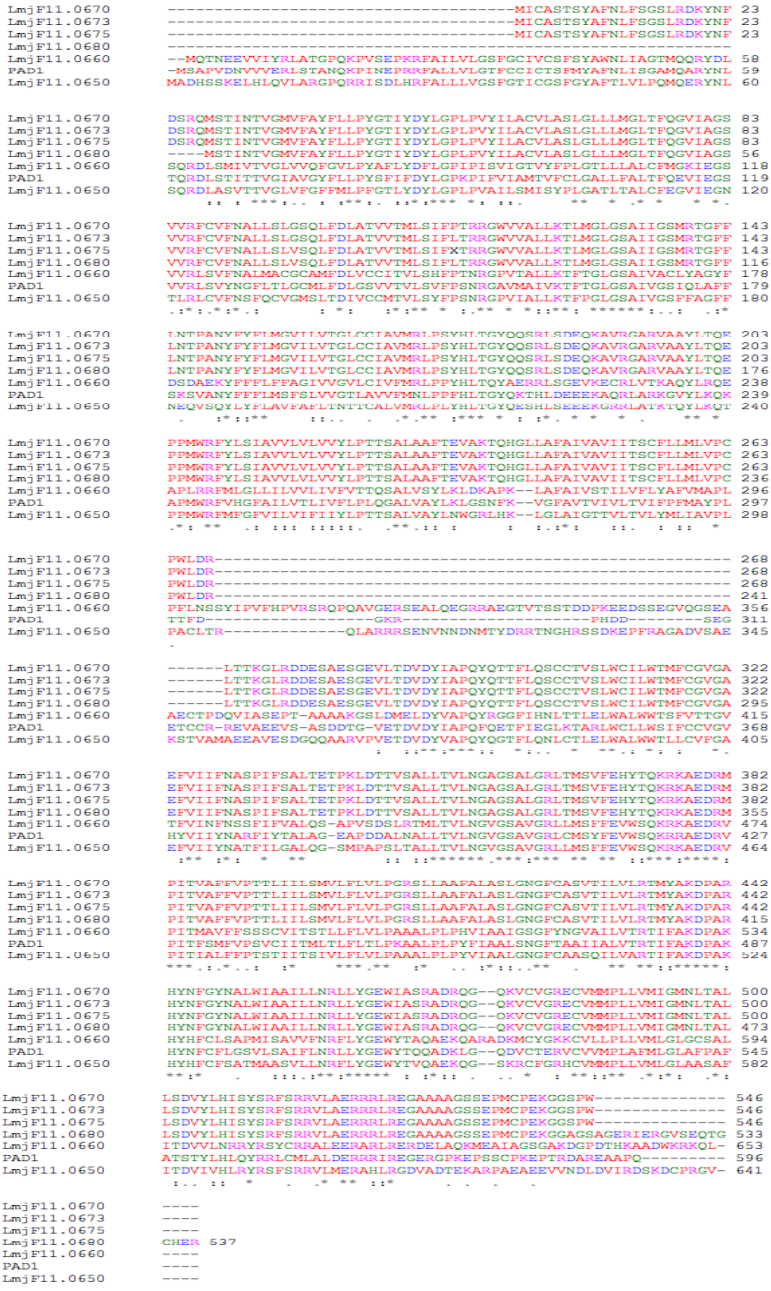


Figure 3.1-4 The closest orthologs to the PAD genes in *L. major* are members of a 6 gene cluster. The *L. major* gene cluster was aligned with PAD1 using clustalW. This gene cluster may represent the *L. major* equivalent of the PAD array, suggesting that the gene duplication event that gave rise to the ancestral PAD genes occurred prior to divergence of the tri-tryp lineage. The asterisks, colons and dots indicate perfect, very high and high conservation between the 5 proteins, respectively.

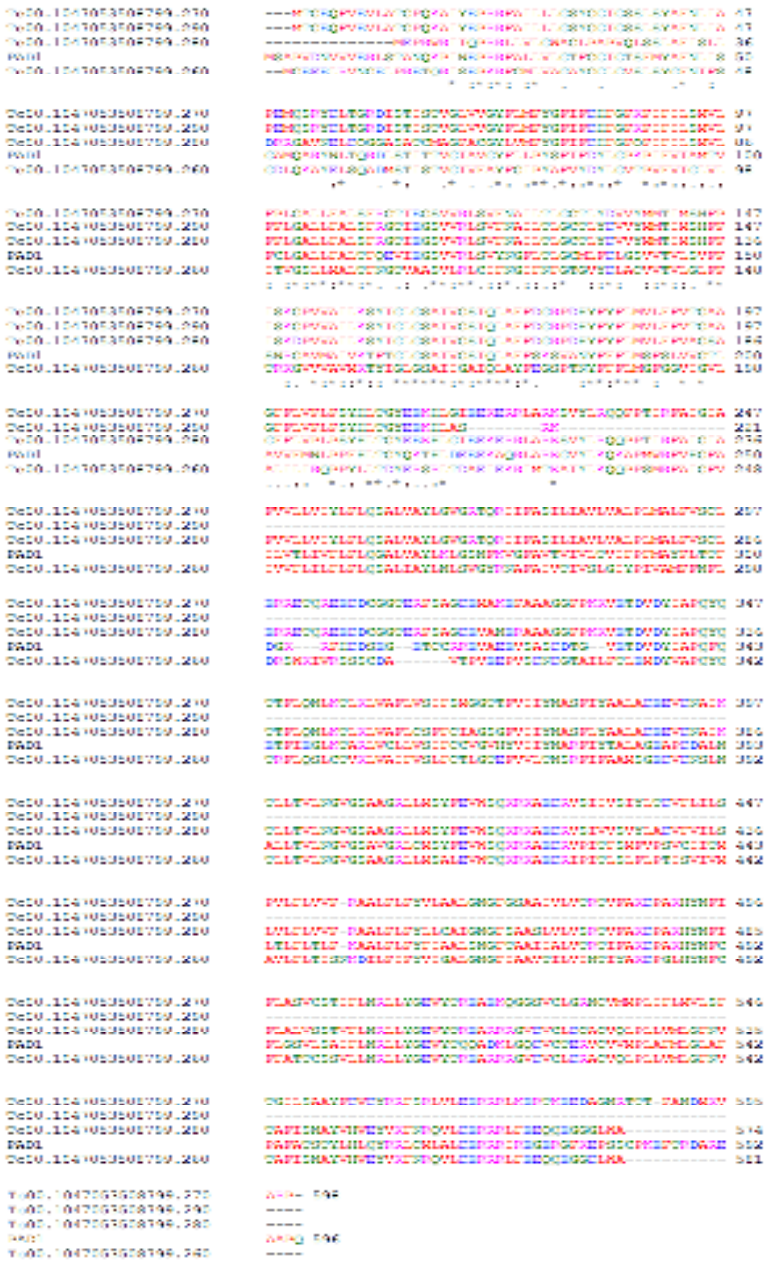


Figure 3.1-5 Close orthologs to the PAD genes in *T. cruzi* are members of a cluster containing at least 4 genes. The second, seventh and thirteenth closest TBLASTX hits to PAD1 in the *T. cruzi* genome are a members of a partially sequenced array of homologous genes containing at least 4 genes. This gene cluster may be part of the *T. cruzi* equivalent of the PAD array. The asterisks, colons and dots indicate perfect, very high and high conservation between the 5 proteins, respectively. Note that Tc00.1047053508799.290 is truncated due to incomplete sequencing.

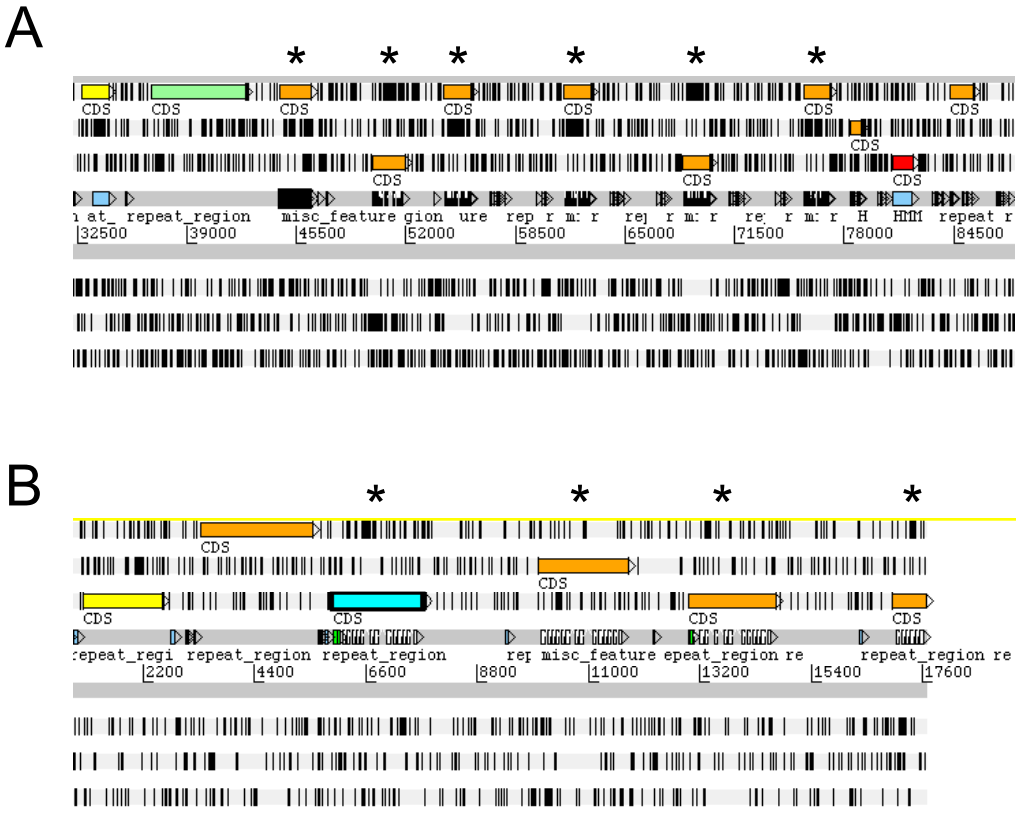


Figure 3.1-6 The *L. major* and *T. cruzi* orthologous PAD arrays. The genomic positions of (A) *L. major* and (B) the partially sequenced *T. cruzi* putative PAD array orthologs were visualised using ARTEMIS (Rutherford *et al.* (2000)). PAD orthologs are indicated with an asterisk. Note that non- orthologous genes annotated as part of this directional gene cluster do not demonstrate homology to genes in the genomic vicinity of the *T. brucei* PAD genes.

3.2 Relationships within the PAD family

The PAD array genes probably arose through several gene duplication events of an original progenitor gene, and thus are highly homologous (Jackson (2007)). A clustalW alignment (Higgins and Sharp (1988)) (Higgins and Sharp (1989)) (Chenna *et al.* (2003)) (Larkin *et al.*(2007)) of all 8 predicted proteins shows 56% overall amino acid identity with PAD1; moreover only 14% of the PAD1 amino acids show no conservation at all with the other PAD members at the corresponding positions (see Figure 3.2-1). A phylogram constructed using a neighbour joining algorithm integrated into the clustalW website (Saitou and Nei (1987)) (Gascuel and Steel (2006)) based upon the protein alignment allows the PAD proteins to be grouped into 3 sets Figure 3.2-1):

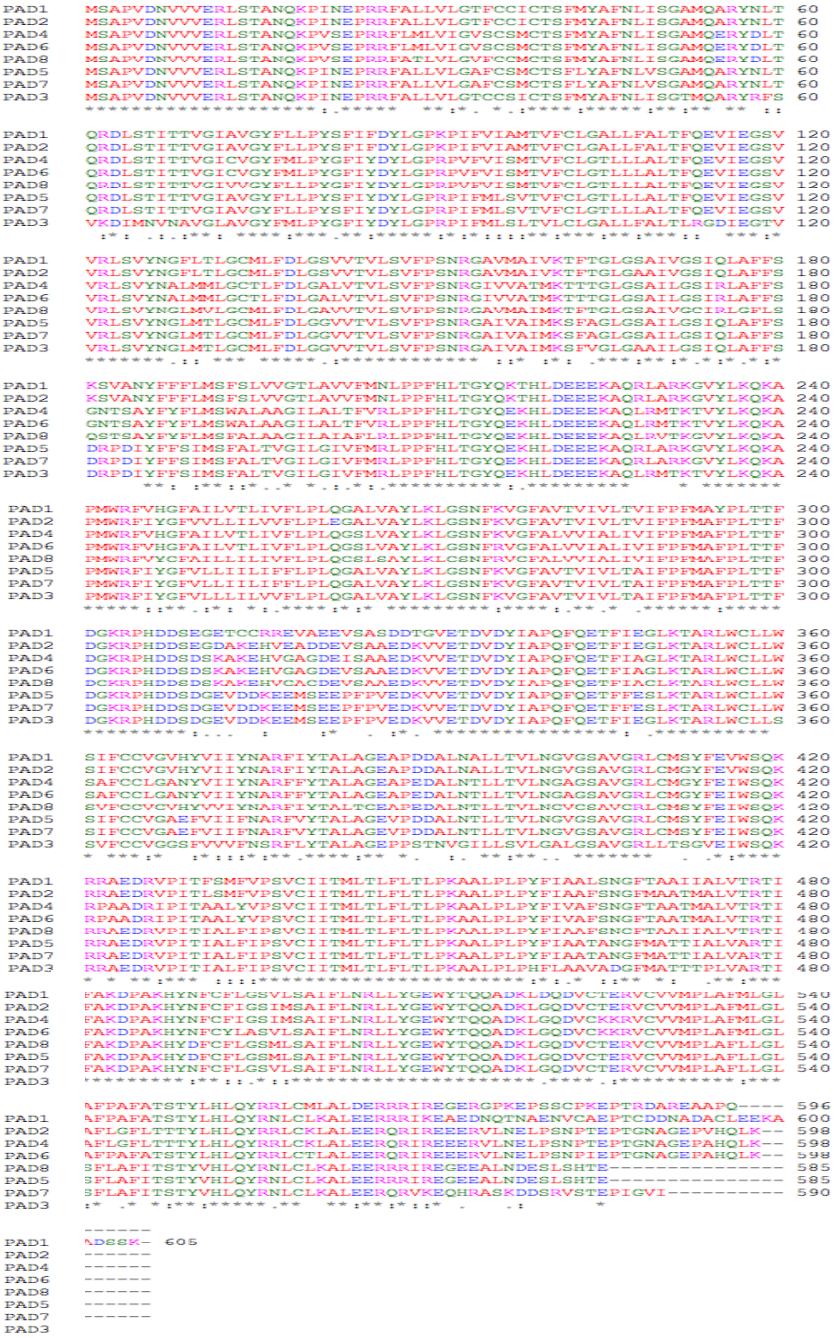
- Group A: PAD1 and PAD2
- Group B: PAD3, PAD5 and PAD7
- Group C: PAD4, PAD6 and PAD8

PAD3 appears to be most distinct from the rest of the array, and PAD5 and PAD7 have identical amino acid and DNA sequences. These trends are reflected in the DNA gene sequence in which a Multalin DNA alignment (Corpet (1988)) shows 66% nucleotide identity between the PAD genes (see Figure 3.2-2).

The PAD genes can also be grouped based upon their position on the chromosome: PAD1 and PAD2 are clustered at the 5' end of the array, after which there is a gap of 4.5 Kb and the other PAD genes follow at more regular intervals of approximately 1.5 kb. This may be indicative of the chromosomal re-arrangements that caused the PAD gene duplication.

It should be noted that, due to the highly repetitive nature of this region of the genome, the database annotation indicates that there may genome mis-assembly and uncertainty in the gene copy number, thus there may be more PAD genes, or in a different arrangement, than is depicted by the GeneDB database.

A



B

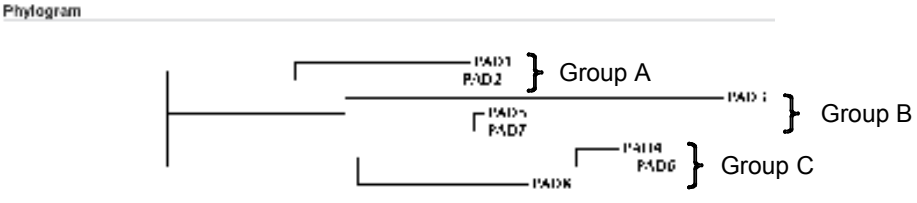
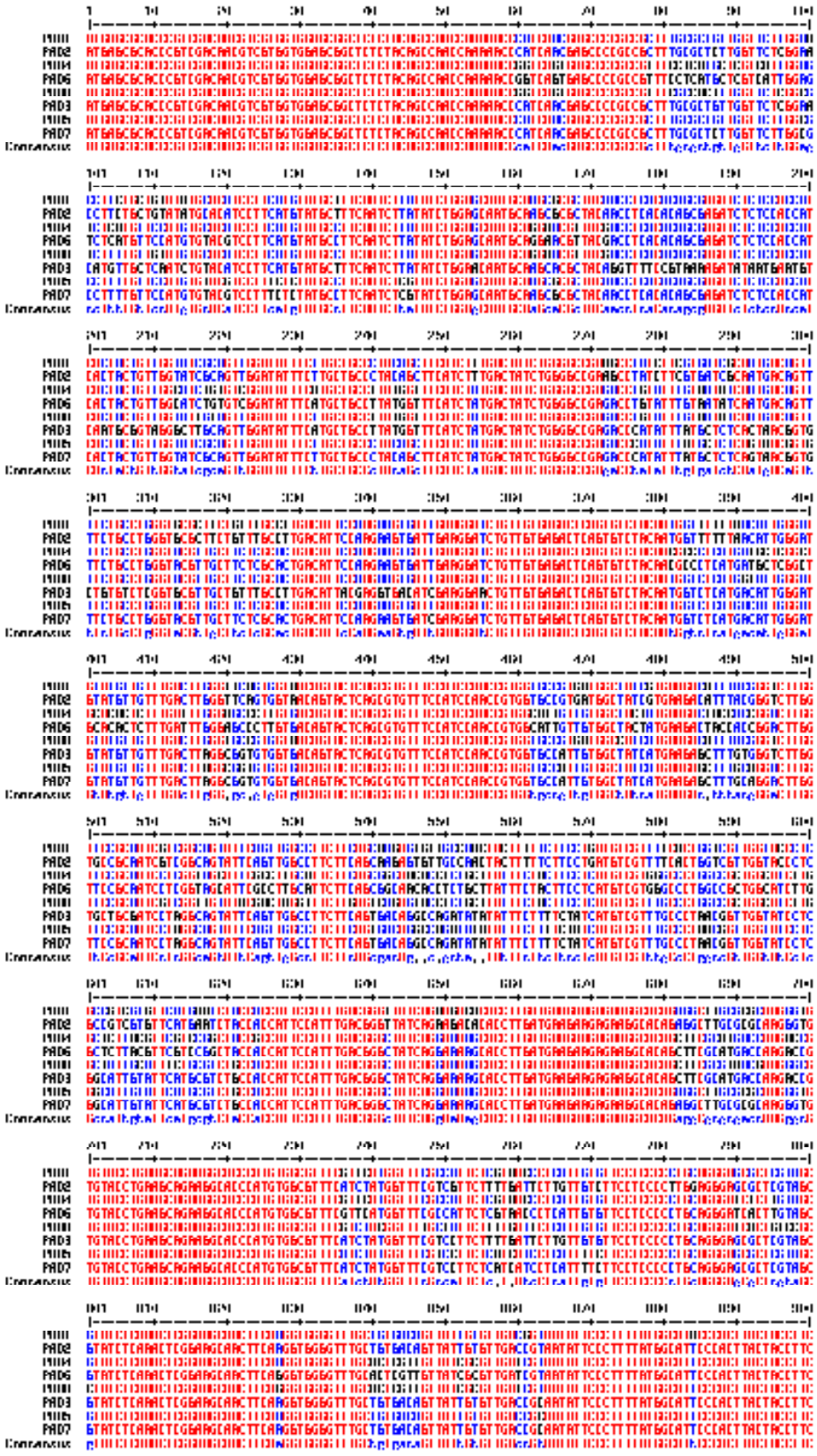
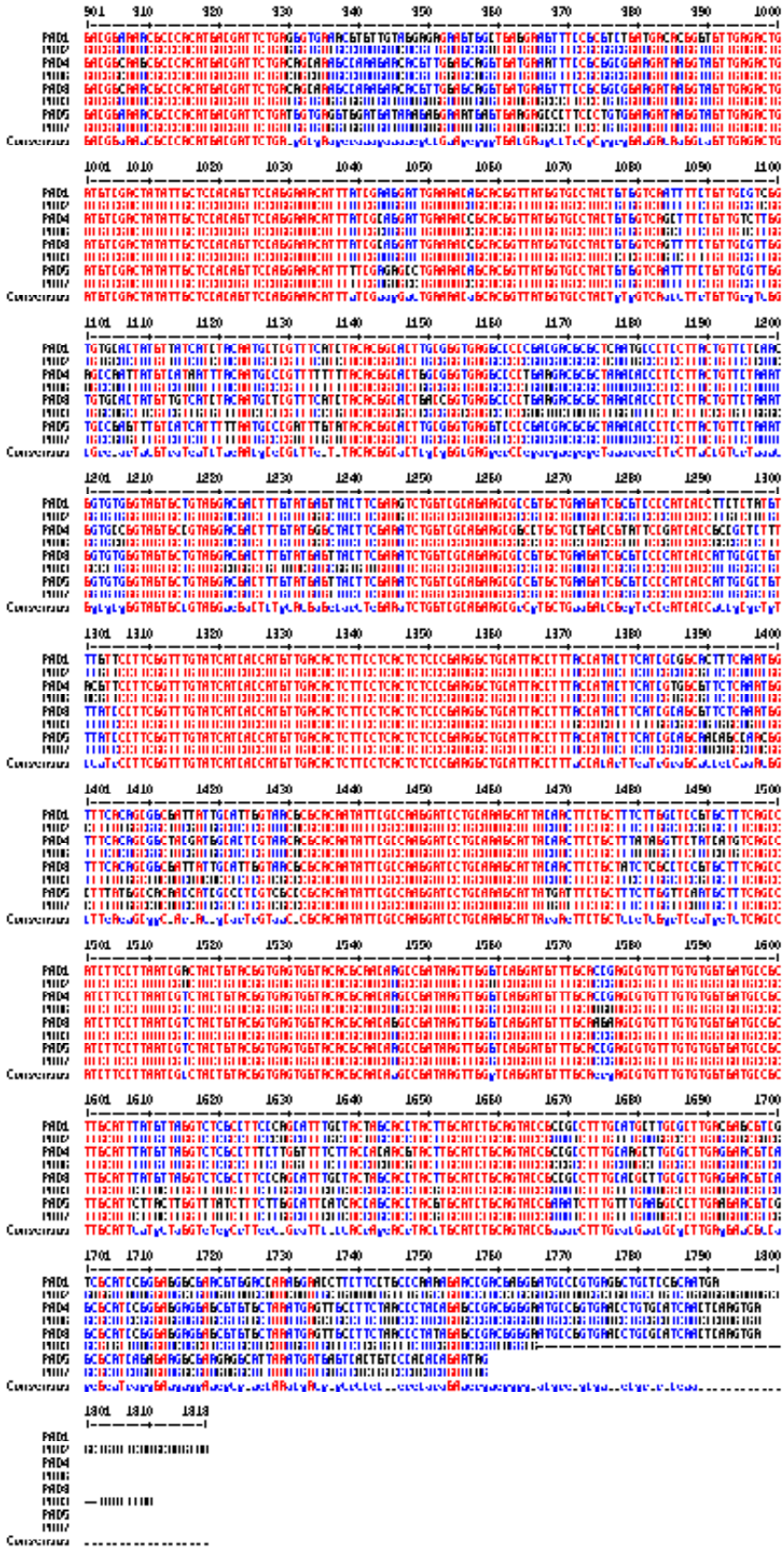

Phylogram
PAD1
PAD2
Group A
PAD4
Group B
PAD3
PAD6
Group C
PAD8

Figure 3.2-1 The relationship between the PAD protein sequences. (A) The PAD protein sequences were aligned using clustalW. 56% of the PAD1 amino acid positions show identity when compared with the same position in the other PAD proteins, and 86% of PAD1 amino acid positions are either identical or show little variation in residue choice. The asterisks, colons and dots indicate perfect, very high and high conservation between the 5 proteins, respectively. (B) The relationship between the PAD protein sequences is depicted in a phylogram constructed using a neighbour joining algorithm. The PAD proteins fall into 3 groups: A: 1 and 2, B: 3, 5 and 7, and C: 4, 6 and 8.

A





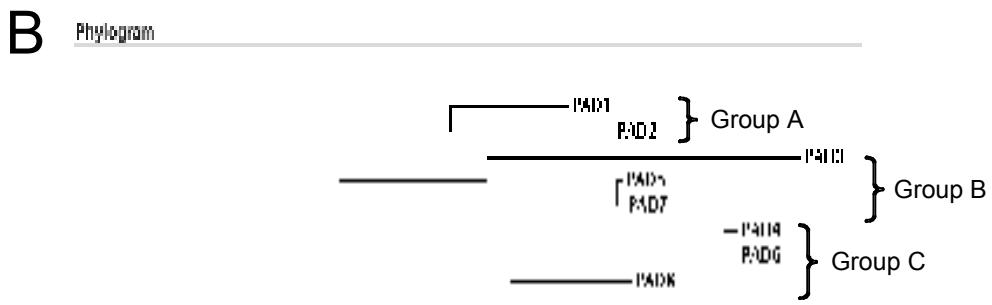


Figure 3.2-2 The relationship between the PAD gene nucleotide sequences. (A) The PAD gene sequences were aligned using Multalin. 66% of the PAD1 nucleotide positions show perfect conservation when compared with the same position in the other 8 PAD genes (indicated by red lettering). (B) A phylogram was constructed using a neighbour joining algorithm. The PAD genes fall into 3 groups: (A) 1 and 2, (B) 3, 5 and 7, and (C) 4, 6 and 8.

3.3 An analysis of the PAD array intergenic regions

The PAD array is approximately 37.1 kb long (including all of the intergenic regions and open reading frames). The average size of the PAD genes is approximately 1.8 kb meaning that approximately 22.8 kb (61%) of the array is intergenic. Thus, any analysis of the PAD array must include the intergenic regions, which are likely to have regulatory functions (Clayton (2002)).

3.3.1 Predicting the positions of the PAD *trans*-splicing sites

In trypanosomatids, mature mRNA transcripts are generated from a pre-cursor polycistronic RNA transcript (Imboden *et al.* (1986)) (Muhich and Boothroyd (1989)) (Rudenko *et al.* (1990)) (Clayton (2002)) through two RNA processing steps: a spliced leader sequence is added onto the 5' UTR of one gene, and a poly-adenosine tail is added to the 3' UTR of the upstream adjacent gene in a linked reaction (LeBowitz *et al.* (1993)) (Ullu *et al.* (1993)) (Matthews *et al.* (1994)). With some exceptions (notably some S phase enriched transcripts in *Crithidia fasciculata* (Mahmood *et al.* (1999)) and genes involved in DNA metabolism in *L. major* (Zick *et al.* (2005))) the 3' UTR is responsible for the stability and translational efficiency of the mature transcript (Clayton *et al* (2002)). Nonetheless, prediction of the *trans* splicing sites is a useful tool for determining which parts of the intergenic region constitute the UTRs of the transcript.

Gopal *et al.* (Gopal *et al.* (2005)) describe development of an algorithm that detects *trans* splicing sites in *L. major*. This algorithm identifies regions of the genome likely to contain splice sites by searching for regions enriched in TC dinucleotides. The splice acceptor AG dinucleotide is then identified by finding the largest inter-AG gap between the AG dinucleotides within the splice site region, and using the 3' AG dinucleotide of this gap as the splice acceptor AG. The advantage of this system is that it is able to predict *trans* splicing signals independently of potentially cryptic start codons.

Similarly, Benz *et al.* (Benz *et al.* (2005)) describe the development of an algorithm that identifies *trans* splice sites within the *Trypanosoma brucei* genome sequence. This algorithm identifies splice leader addition sites by finding a polypyrimidine tract of at least eight pyrimidines immediately upstream of a predicted open reading frame, and then determining the closest AG dinucleotide downstream of that tract. The linked poly-adenosine tail addition was predicted to occur at one or more adenosines between 80 and 140 bases upstream of this polypyrimidine tract. Thus, the algorithm defines the location of the poly-adenosine tail addition by finding the adenosine closest to the position 72 nucleotides upstream of the polypyrimidine tract. This approach relies on the identification of the correct start codon for an open reading frame, and may thus incorporate errors from mis-annotation of the genome database. Nonetheless, this provides an effective way of estimating most *trans* splicing sites in *Trypanosoma brucei*.

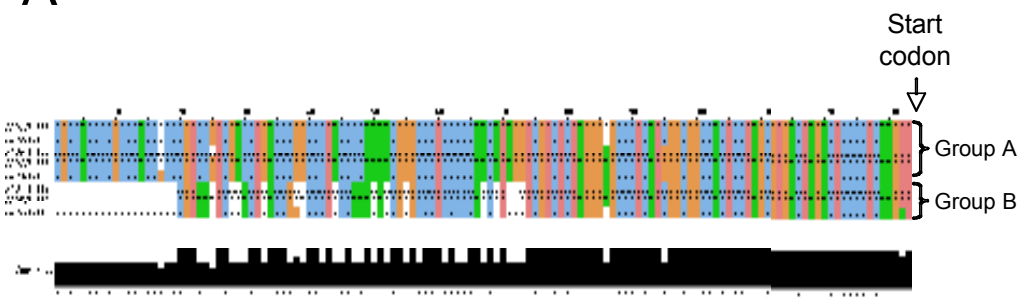
The genome annotation performed by Benz *et al*. using this approach was downloaded (http://web.cgb.ki.se/daniel/splicemodel.php) and used to estimate the 5' and 3' UTRs of the PAD genes for subsequent analysis.

3.3.2 Aligning the PAD UTRs

The predicted PAD 5' UTRs were aligned using clustalW (Higgins and Sharp (1988)) (Higgins and Sharp (1989)) (Higgins *et al.* (1994)) (Lopez and Lloyd (1997)) (Larkin *et al.*(2007)) and visualised using Jalview (Clamp *et al.* (2004)). As expected from an array that probably arose through tandem duplication, the PAD 5' UTR sequences are very similar (see Figure 3.3-1). They are either 112 or 133 nucleotides long, which is longer than the median of 68 nucleotides found by cDNA analysis (Benz *et al.* (2005)) but within the range of 40 – 200 nucleotides found by surveying the Genbank database (Gopal *et al.* (2005)). The 86 nucleotides proximal to the start codon are almost identical between the 5' UTRs and only differ at two positions. The 5' UTRs fall into 2 distinct groups (see Figure 3.3-1):

- Group A: PAD4, PAD5 and PAD7
- Group B: PAD1, PAD2, PAD3, PAD6 and PAD8

A



B

	5' UTR	3' UTR
PAD1	133	754
PAD2	133	4276
PAD3	133	1374
PAD4	112	1288
PAD5	112	1017
PAD6	133	1282
PAD7	112	1005
PAD8	133	2517

Figure 3.3-1 The positions and alignments of the predicted PAD UTRs. (A) The PAD array predicted 5' UTR nucleotide sequences were aligned using clustalW and visualised using Jalview. The 5' UTRs appear to fall into 2 groups: A: PAD4, 5 and 7, and B: PAD1, 2, 3, 6 and 8. (B) The length of the predicted PAD 5' and 3' UTRs is shown. The PAD 5' UTRs are either 112 or 133 nucleotides in length whereas the PAD 3' UTRs are more heterogeneous and range from 754 to 4276 nucleotides in length; this heterogeneity appears improbable, and may in fact represent the inherent difficulties in predicting RNA processing signals.

The predicted PAD 3'UTRs fall into two approximate groups based on their sequence alignment (see Figure 3.3-2):

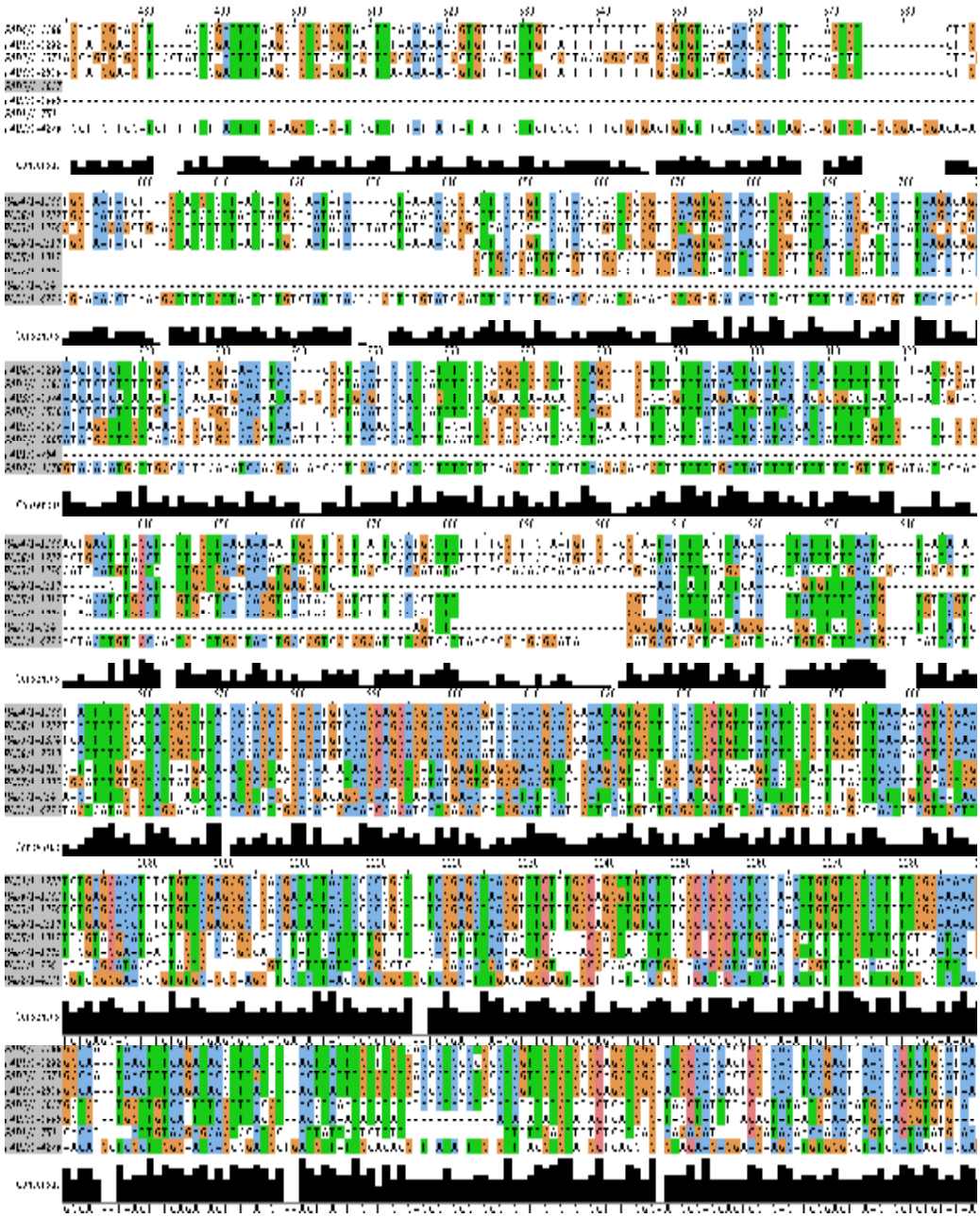
- Group A: PAD1, 5 and 7;
- Group B: PAD4 and 6

PAD2, 3 and 8 appear to form distinct groups and vary in their proximity to group A and B.

However, the extreme heterogeneity in the size of the PAD 3' UTRs appears unlikely, especially when considering the PAD genes' otherwise homologous nature. For example, the predicted PAD2 3' UTR is more than 5 times the length of the predicted PAD1 3' UTR; it would appear more likely that the UTR predictive algorithm does not predict RNA processing sites accurately in longer intergenic regions. Moreover, this size heterogeneity was not detected when PAD1, 2, 5 and 7 mRNA transcripts were analysed by Northern blot (see Section 5.1). It is likely that in intergenic regions the predictions are inaccurate and underestimate the length of non-UTR intergenic regions.

Nonetheless, these groups may be reflective of expression profile or translational efficiency; one might speculate that there may be functional redundancy between the proteins expressed within these groups.

A



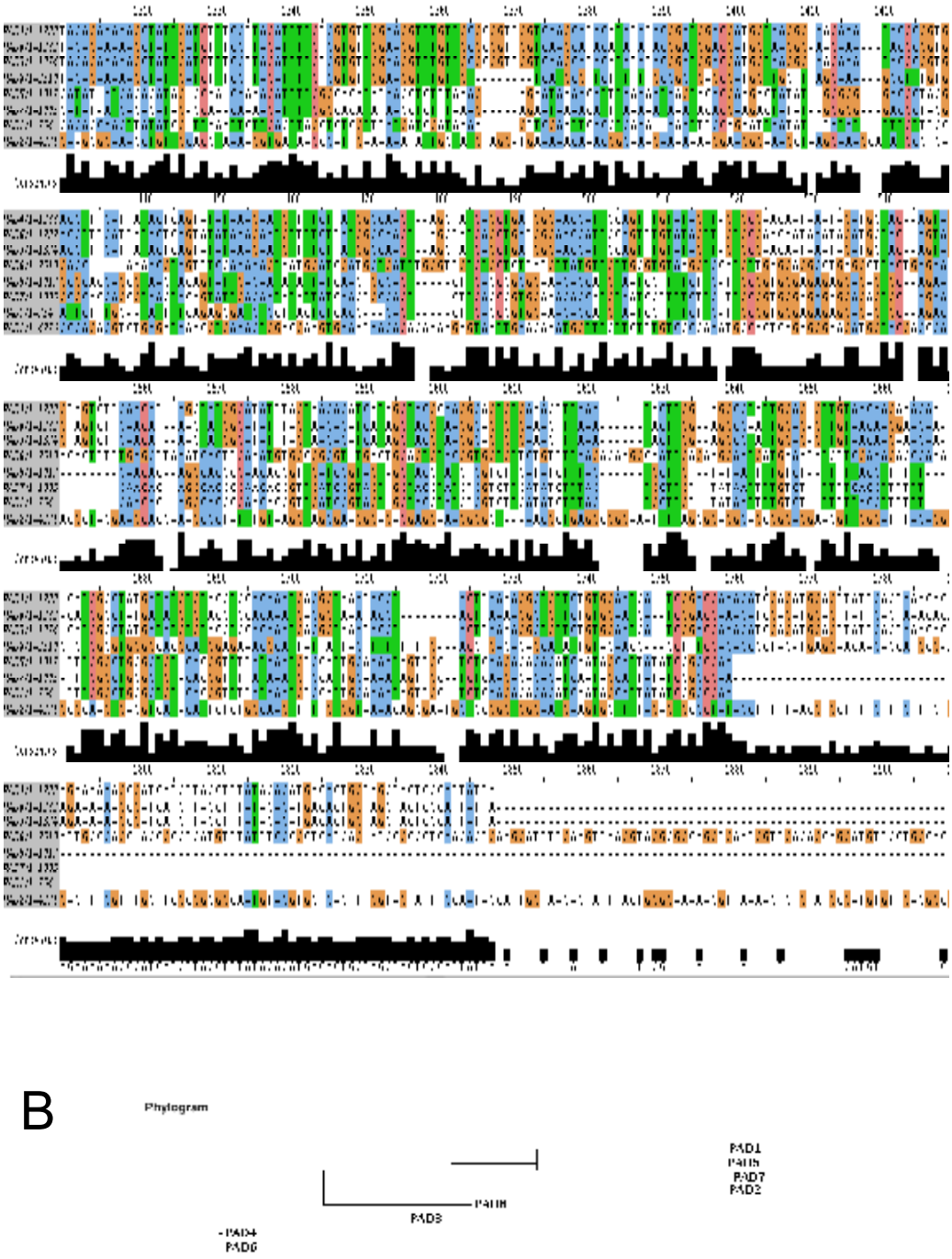


Figure 3.3-2 A comparison of the predicted PAD 3' UTRs. (A) The predicted 3' UTRs of the PAD genes were aligned using clustalW and visualised using Jalview. Note that the PAD2 and 8 predicted 3' UTRs were truncated to save space. (B) A phylogram was constructed based upon the alignment using a neighbour joining algorithm. PAD1, 5 and 7 and PAD4 and 6 form distinct groups. PAD3, 8 and 2 are less closely related to other genes within the array.

3.3.3 An analysis of the repetitive elements within the PAD intergenic regions

A multiple alignment analysis of the PAD intergenic regions (see Appendix A) as designated by the GeneDB database identified 6 elements (designated I to VI) that are repeated through the intergenic regions. These range from 59 to 955 nucleotides in length (see Figure 3.3-3) and are repeated at variable frequencies throughout the array.

Element I is present in the intergenic region upstream of PAD1 (designated UP-1), as well as intergenic regions between PAD1-2, PAD2-3, PAD5-6, PAD7-8, but was not predicted to form part of any of the PAD 5' or 3' UTRs. Thus element I is therefore either non-functional or has a role directing the 5' *trans*-splicing reaction.

Element II contains a predicted 5' *trans*-splicing signals, whereas element IV contains a predicted 3' poly-adenylation signal. Element III appears to contain both 5' *trans* splicing signal and 3' poly-adenylation signals. These elements may also have roles in the stability or translational efficiency of the resulting mature mRNA transcript.

PAD4, PAD6 and PAD8 share elements II and III and may thus be co-regulated. PAD2 is unique in that it contains no repeated elements in the 3' UTR; the PAD2 transcript was shown to be elevated in procyclic forms (see Section 5.1) thus this non-repeated element presumably contains a procyclic form stabilisation element.

Mayho *et al.* (Mayho *et al.* (2006)) identified an 8 nucleotide U-rich element (uauuuuuu) that was over-represented in procyclic form-specific gene 3' UTRs. This element was found to be present in 31% of procyclic enriched transcripts compared with just 9% of blood form enriched transcripts and at an overall frequency of 9.6%. An analysis of the repetitive intergenic elements of the PAD array found this sequence twice in element IV and once in element V. Additionally, it was found that this sequence was present in the non-repetitive regions of the PAD2 and PAD3 3' UTR; thus, it appears that this element is present in the 3' UTR of all of the PAD

genes except for PAD1. However, a Northern analysis to dissect the expression of the different PAD genes through the life cycle revealed that PAD5 and 7 are not expressed in the procyclic form of the parasite (see Section 5.1), despite containing this U-rich sequence. We cannot therefore make any inference from the presence of this element in the PAD 3' UTRs.

Interestingly, element VI is present in the 3' UTR region of PAD1, PAD5 and PAD7, all of which are stumpy form specific mRNA transcripts (see Section 5.1) and this region may therefore contain a stumpy-specific stabilisation element. Element V is only present in the PAD5 and PAD7 3' UTRs and is not found in the PAD1 3' UTR. This element may have an additional role unrelated to transcript stabilisation in stumpy form cells or may simply represent a non-regulatory sequence that was duplicated along with PAD5 and 7.

Given that PAD1, 5 and 7 are stumpy specific transcripts (see Section 5.1), one might speculate that the elements responsible for stability in stumpy form cells might allow identification of other stumpy specific transcripts within the *T. brucei* genome. To this end, elements V and VI were used as a query sequence in a BLASTN search of the *T. brucei* genome. However, other than PAD array sequence, no significant hits were returned from this search. Moreover, the degree of homology shown by the identified hits was low, and the position on the chromosome was often on the non-coding DNA strand.

A

Intergenic element	*Length (bp)*	*PAD array intergenic regions*								
		UP-1	*1-2*	*2-3*	*3-4*	*4-5*	*5-6*	*6-7*	*7-8*	*8-DOWN*
I (orange)	132	+	+	+			+		+	
II (light blue)	59	+	+	+	+	+	+	+	+	
III (green)	955				+	+		+		+
IV (dark blue)	432					+		+		+
V (purple)	766						+		+	
VI (red)	272		+				+		+	

B

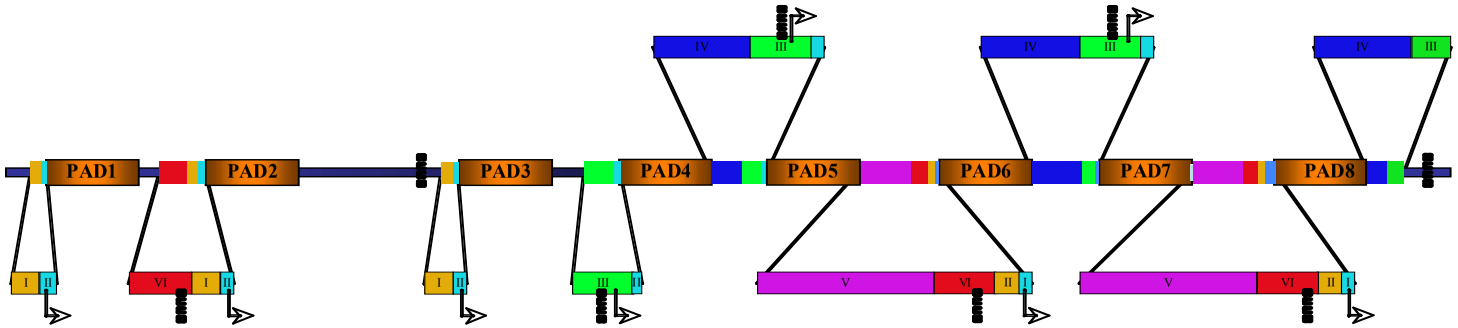


Figure 3.3-3 Analysis of the repetitive elements within the PAD array intergenic regions. (A) The 6 intergenic elements that are repeated through the PAD array are represented in a table. UP-1 refers to the intergenic region upstream of PAD1, 8-DOWN corresponds to the intergenic region downstream of PAD8. The other intergenic regions are indicated by their flanking genes (e.g. 4-5 corresponds to the intergenic region between PAD4 and 5). "+" means that the element is present within the corresponding intergenic region. (B) The elements and their positions are represented in the diagram. Note that element 3 (green) is truncated downstream of PAD8. 5' UTR trans-splicing sites are depicted as arrows, and 3' UTR splicing sites are depicted as dashed lines.

3.3.4 Predicting the secondary structure of the PAD 3'UTRs

The structure of 3'UTRs is likely to be important for RNA binding protein specificity and therefore has an essential role in transcript stability and translational efficiency (Hotz *et al.* (1997)) (Irmer and Clayton (2001)). Particularly, it is thought that mRNA regulatory proteins bind the single stranded RNA and hairpin loop region of folded mRNA sequence (Hehl *et al.* (1994)) (Furger *et al.* (1997)) (Hotz *et al*. (1997)) (Mayho *et al*. (2006)). Experimental elucidation of RNA structure is expensive, technically challenging and time consuming, and currently there are no bioinformatics programs available that predict the three-dimensional structure without some experimental data and user defined parameters.

However, there are, several programs available that predict the secondary structure of RNA molecules based on its primary sequence. These programs operate by calculating the free energy associated with a structure based on the base pairing of the component nucleotides. Using a greedy algorithm, the structure is then modified until the minimum free energy (MFE) structure is found.

However, secondary structure prediction can be complicated by:

- The presence of solutes and ions that might affect RNA folding
- A variety of different RNA structures within a population
- More complex structures such as pseudoknots that are often badly predicted
- Interactions with RNA binding proteins
- Interactions with other RNA molecules
- Variations in temperature

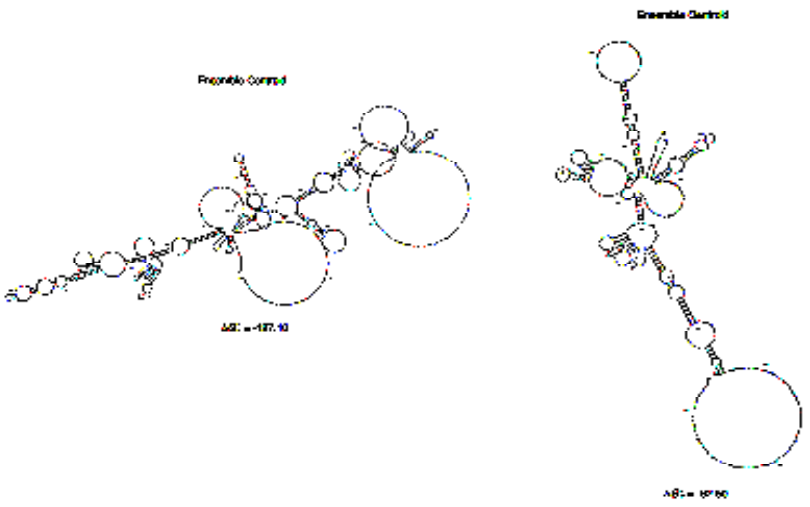
For these reasons, the MFE structure may not be the biologically relevant structure and often a number of sub-optimal free energy structures are generated for comparison. Although examining sub-optimal structures can be valuable, when several RNA molecules are being compared the volume of data returned is

impractical, and often a single consensus structure is more appropriate. In the case of Sfold (Ding and Lawrence (1999)) (Ding and Lawrence (2001)) (Ding *et al.* (2004)), the sub-optimal structures are statistically sampled and a consensus, or centroid, structure is generated based on minimal distance from the sampled sub-optimal structures. Sfold was therefore used in preference to other RNA secondary structure prediction programmes, such as Mfold.

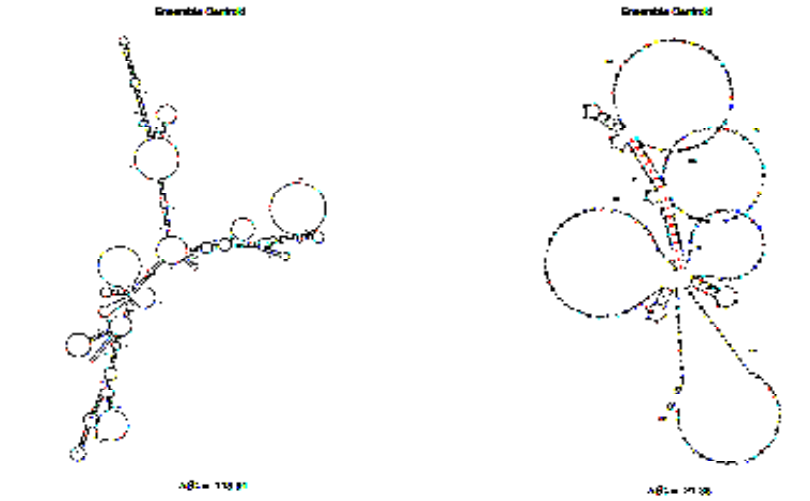
Given that the elements described in Section 3.3.3 are present in several different contexts (e.g. in the PAD3 3' UTR element III is downstream of a non-repetitive region, whereas in the PAD4 3' UTR it is downstream of element IV), one would predict that they are able to fold independently of adjacent nucleotide sequences. Elements IV and V were submitted to Sfold in their entirety, and elements III and VI were modified prior to submission such that only sequence present in the 3' UTR were analysed.

Figure 3.3-4 shows that element III is predicted to have two large single stranded loops extending from a longer stem, and 3 smaller loops. Element IV has a large terminal single stranded loop at the distal end from a more complex group of small stem-loops. Element V has a large branch causing it to form a Y shaped structure. Additionally there are 3 principal loops and a number of smaller loops interspaced through the structure. Element VI has the largest proportion of single stranded regions of the four elements. Only 23 (9%) nucleotides from a total of 251 form bonds with other nucleotides in the structure suggesting that this element is extremely unstable. This element consists primarily of 5 large and 2 small single stranded loops which may provide access for RNA binding proteins to form complexes and regulate the stability of the rest of the transcript.

Caution should be exercised when interpreting structural predictions in the absence of experimental data. Nonetheless, this kind of analysis can be useful in identifying regions that might be targeted in a mutation analysis of PAD gene regulation.



III IV



V VI

Figure 3.3-4 The predicted secondary structure of elements within the PAD 3' UTR as predicted by Sfold. The elements contained within the PAD 3' UTRs were submitted for secondary structure prediction analysis using Sfold. Where an element contained a poly-adenylation addition site only the sequence that formed part of the predicted 3' UTR was submitted.

3.4 Detecting conserved domains within PAD proteins

There are numerous programmes available that can help to characterise the function of a protein by detection of domains that have an associated function. Several of these were used to analyse the PAD proteins.

3.4.1 Pfam analysis

The Pfam database (Sonnhammer *et al.* (1998)) (Finn *et al*. (2006)) is a large collection of protein families and conserved domains that can be interrogated with a query sequence to help characterise a protein. Pfam consists of 9318 protein families (http://pfam.sanger.ac.uk/) that are divided into the high quality, manually curated PfamA and the automatically generated PfamB (Sonnhammer *et al.* (1997)) (Finn *et al*. (2006)).

Each family in Pfam is represented by a multiple sequence alignments that gives rise to a hidden Markov model. A hidden Markov model is a statistical model that describes the 'hidden' parameters of a stochastic process. In this example, the hidden Markov model describes the mathematical relationship between proteins within the same family, and the probability of a query protein being a member of the same family is calculated.

Both the manually edited PfamA and the automatically generated PfamB components were searched using the PAD proteins as a query sequence. Thus, all of the PAD proteins contain a clear Major Facilitator Superfamily (MFS) 1 domain (Pfam accession number IPR011701) and as such are members of this clan (see Table 3.5-1). Members of the MFS typically have 6, 12, 14 or 24 transmembrane helices and transport a substrate by facilitating an existing chemiosmotic gradient (Rubin *et al.* (1990)) (Paulsen *et al*. (1998)). Transport is not active as in the ATP binding cassette (ABC) transporters and can occur bidirectionally through uniport, symport or antiport (see Figure 3.4-1 and Section 1.7.3). No other domains were detected within the primary sequence using Pfam.

3.4.2 CDD analysis

PAD1 was also analysed using the Conserved Domain Database (CDD). CDD is a collection of ancient domains and full length proteins that have been aligned and a profile has been constructed using a reverse position-specific BLAST algorithm. A query sequence is then compared with the position score matrices prepared from the underlying domain alignments. When PAD proteins were used as a query sequence it was found that they have a nodulin-like domain (Pfam06813) (see Table 3.5-1), a conserved region within plant nodulin-like proteins that is often found in MFS proteins. Nodulin-like transporters are largely uncharacterised, but several are known to transport nitrate (Vincill *et al.* (2005)) and other small solutes.

3.4.3 Prosite analysis

Prosite (Hulo *et al*. (2006)) is a collection of protein domains, families, and functional sites that have been collected into groups based upon their sequence similarity and likely function. By analysing the constant and variable regions within these groups a 'signature' for each group has been built that allows members to be distinguished from unrelated proteins. A query sequence is then analysed and compared with the database of signatures to determine which, if any, of the groups it belongs to. Using the PAD proteins as query sequences no matches were found (see Table 3.5-1). The most likely reason for this is that functional analyses of transporter molecules have largely failed to identify the specific characteristics required for substrate specificity mechanism of action.

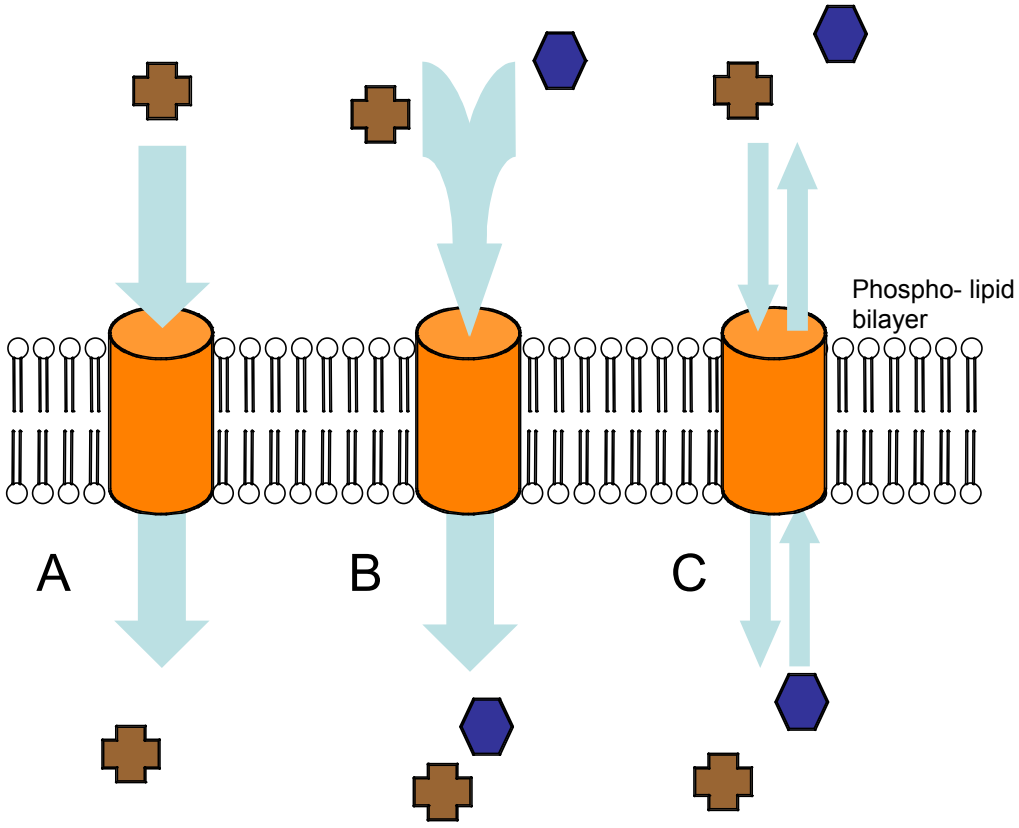


Figure 3.4-1 A schematic showing the mechanisms of transport for major facilitator superfamily transporters. (A) Uniport where a single substrate is transported across the membrane, (B) Symport where two substrates are transported in the same direction across the membrane and (C) Antiport, where two substrates are transported in opposite directions across the membrane. The blue and brown shapes depict the transported molecules, the arrows show the direction of transport and the orange tubes the MFS transporters.

3.5 Predicting the localisation of the PAD proteins

Determining the cellular localisation of a protein is a vital step to elucidating its function, and there are a number of programmes available that can to identify cellular targeting sequences.

SignalP3.0 (Neilsen (1997)) (Bendtsen (2004)) (Emanuelsson (2007)) uses a combination of neural networks and hidden Markov models to predict the presence and location of a signal peptide within the query sequence. Signal peptides are amino terminal peptides approximately 15 – 30 residues long that are responsible for insertion of the protein into the membrane during translation, and target the protein for translocation across the plasma membrane in prokaryotes, and across the endoplasmic reticulum in eukaryotes. Typically, a signal peptide contains a positively charged region at the amino terminus (n-region) followed by a hydrophobic region (h-region) of a least 6 residues and ending in a carboxyl terminal region containing polar, uncharged residues (c-region). A signal peptide is generally cleaved and, unless it contains a retention sequence, the protein is secreted by default in secretory vesicles. When the PAD proteins were used as a query sequence in the SignalP programme, no signal peptides were detected, but a signal anchor, or uncleaved signal peptide, was predicted with a strong probability (see Table 3.5-1). Signal anchors have an almost identical structure to signal peptides, but differ in that they do not contain a cleavage site and therefore remain part of the mature protein. Proteins containing signal peptides where the amino terminus is cytoplasmic are termed type II membrane proteins, as is the case with the PAD proteins.

WolfPSort (Horton *et al.* (2006)) (Horton *et al.* (2007)) analyses a sequence to detect subcellular localisation amino acid sequences that may elucidate a protein's final position within the cell and is most able to detect signals that localise proteins to the cytosol, nucleus or the mitochondrion. WolfPSort is designed to analyse animal, plant or fungal proteins rather than protozoan proteins, so all of these options were tried sequentially. No localisation motifs were detected in the PAD proteins; in particular, it is notable that targeting motifs for the membrane bound organelles such

as the mitochondrion were not found. A possible vacuolar targeting motif (TPLK) was detected indicating potential endosomal trafficking.

TargetP (Emanuelsson (2000)) (Nielsen (1997)) analyses a protein to detect either mitochondrial targeting motifs or signal peptides. None were found in the PAD proteins, suggesting that the protein is not targeted to the mitochondrion and is not cleaved and secreted.

Table 3.5-1 A summary of the PAD protein bioinformatics data

Programme	Detection	Results	Reliability
Pfam	Conserved domains using the Pfam database	MFS-1	E = 1.7E^{-6}
CDD	Conserved domains (a.a. 44 - 217)	Nodulin-like domain	E = 4E^{-6}
Prosite	Conserved domains	None found	
Prosite	Conserved domains	None found	
WolfPSort	ER retention motif in the C-terminus	None found	
	ER Membrane Retention Signals	None found	
	SKL: peroxisomal targeting signal in C-terminus	None found	
	2nd peroxisomal targeting signal	None found	
	Possible vacuolar targeting motif	TPLK found at pos. 449	
	RNA-binding motif	None found	
	Actinin-type actin-binding motifs	None found	
	N-myristoylation pattern:	None found	
	Farnesylation/Geranylgeranylation motif	None found	
	Transport motif from cell surface to Golgi	None found	
	Tyrosines in the tail	None found	
	Dileucine motif in the tail	None found	
big-PI	GPI anchor cleavage sites	None found	
PSORT	Signal peptide	None found	
	Mitochondrial localising peptide	None found	
SignalP3.0	Signal peptide	None found	
	Anchor peptide	Yes	P = 0.836

3.6 Predicting the secondary structure of the PAD proteins

The web based topology prediction programme TMHMM (Tusnády and Simon (1998)) (Tusnády and Simon (2001)) was used to interrogate the amino acid sequence of the PAD proteins. TMHMM uses a hidden Markov model to analyse the differences in amino acid composition within the query sequence (rather than utilising the absolute frequency of individual amino acids), such that the positions of the transmembrane spanning (TMS) regions, intracellular loops and extra-cellular loops are predicted. The data from HMMTOP was visualised as a secondary structure using the downloadable programme TMRPres2D (Spyropoulos *et al.* (2004)). In total, 14 TMS were predicted for the PAD proteins (see Figure 3.6-1A and Appendix A), consistent with them being part of the major facilitator superfamily and the Pfam data. In addition, the positions of two large intra-cellular loops are indicated for PAD1, and both the amino terminus and the carboxyl terminus are predicted to be intracellular.

Charged amino acids are thought to be important in substrate binding, particularly if present in membrane helices (Guan and Nakae (2001)) (Shelden *et al.* (2003)) (Manoharan *et al.* (2006)) (Valdés *et al.* (2006)). Rahman *et al.* (Rahman *et al*. (1999)) showed that charged amino acid residues in helices 7 and 10 are essential for substrate binding in the rat monocarboxylate transporter MCT1 (see Section 3.8 for further discussion of PAD protein similarities with carboxylate transporters). Interestingly, only PAD 2, 4 and 6 have charged residues in helix 7, while PAD1, 5, 7 and 8 appear to have a reciprocal charged residue in helix 9, and PAD 3 in helix 13 (see Figure 3.6-1); it would be informative to determine whether the PAD proteins divide into these three groups in either substrate specificity, affinity or translocation.

In contrast, all of the PAD members contain a charged residue in helix 4, 5 and 13, suggesting that these residues are essential for substrate binding or translocation; any mutagenesis strategy to determine which parts of the polypeptide are important for function would focus on these residues.

There are examples that show there is the potential for overlap of function between transporter and receptor molecules. Özcan *et al*. (Özcan *et al*. (1998)) show that two proteins of the major facilitator superfamily in *Saccharomyces cerevisiae* are able to sense the presence of glucose and transmit the signal without substrate translocation. These proteins have long (218 and 314 residues) carboxyl terminal cytoplasmic domains containing a 25 amino acid residue motif that is necessary for signal transduction. In contrast, the PAD genes have cytoplasmic carboxyl domains of less than 50 residues and do not contain the glucose sensing motif, suggesting they function as conventional transporters.

A.

	PAD 1	PAD 2	PAD 3	PAD 4	PAD 5/7	PAD 6	PAD 8
Helix 1	27-46	27-46	27-46	27-46	27-46	27-46	27-46
Helix 2	67-85	67-86	65-84	67-85	65-84	67-85	65-84
Helix 3	92-111	93-112	93-112	92-111	93-112	92-111	93-112
Helix 4	130-149	129-148	129-148	130-149	129-148	130-149	129-148
Helix 5	162-179	155-174	155-174	162-181	155-174	162-181	155-174
Helix 6	188-205	183-205	185-204	186-205	185-204	186-205	187-205
Helix 7	245-268	245-268	245-268	240-259	245-268	240-259	240-259
Helix 8	277-300	277-295	277-295	280-299	277-295	280-299	280-299
Helix 9	355-374	355-374	355-374	355-374	355-374	355-374	355-374
Helix 10	395-414	395-414	395-414	395-414	395-414	395-414	395-414
Helix 11	427-450	427-450	427-447	427-448	427-450	427-448	427-450
Helix 12	459-476	459-477	460-477	457-476	459-477	457-476	459-477
Helix 13	489-507	490-509	490-509	489-507	490-509	489-507	490-509
Helix 14	528-547	528-547	528-547	528-547	528-547	528-547	528-547

B.

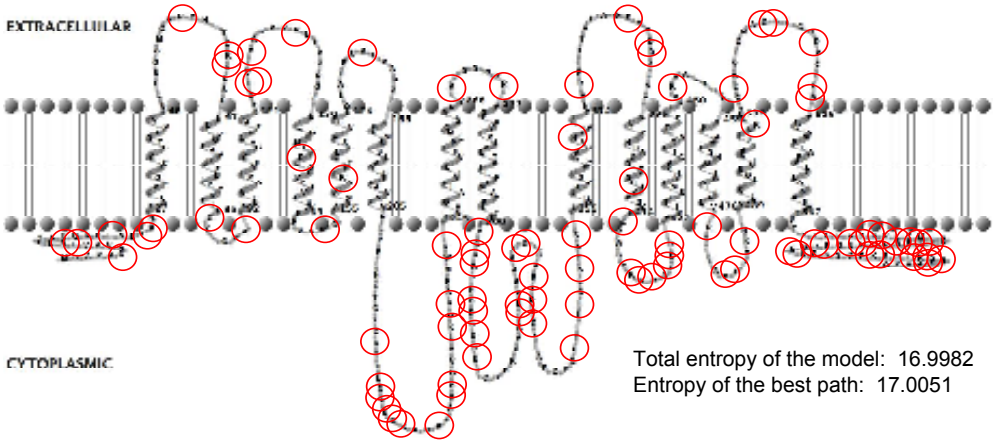


Figure 3.6-1 The secondary structure prediction of the PAD proteins. (A) The positions of the transmembrane spanning regions as predicted by HMMTOP are shown for each of the PAD proteins. Helices containing charged amino acid residues are circled in red. (B) The secondary structure of PAD1 was visualised as a representative of the PAD cluster using TMRPres2D based upon the HMMTOP output. Charged amino acid residues are circled in red. The secondary structure for the other PAD proteins can be found in appendix A.

3.7 Predicting the three-dimensional structure of the PAD proteins

Predicting the three dimensional structure from the amino acid sequence of a protein is inherently difficult and prone to error. Programmes that do this rely on finding homology to experimentally verified structures and use this as a template with which to predict the structure of a query sequence. This is particularly challenging with regard to transmembrane proteins as the membrane forms an integral part of the structure and can make predictions problematic. Several programmes (e.g. 3D jigsaw (Bates and Sternberg (1999)) (Bates *et al.* (2001)) (Contreras-Moreira *et al*. (2002))) were used to predict the structure of PAD proteins but were unsuccessful due to the lack of an existing experimentally verified structure to use as a template.

3.8 Using PSI-BLAST to help elucidate the PAD proteins' function

Finally, to identify homologues that may have an identified substrate or may otherwise help elucidate PAD function a Position Specific Iterated Basic Local Alignment Search Tool (PSI-BLAST) search was performed. PSI-BLAST performs a BLASTP search that identifies homologues in a database of sequences. Using these 'hits' a profile is built a profile using a hidden Markov model such that frequently occurring amino acids at specific positions are given more importance; the BLAST search is then re-iterated using this profile. PSI-BLAST is more sensitive than conventional BLAST and may detect homologues that would otherwise go unidentified. Furthermore, the proteins identified through BLAST based algorithms can be grouped based upon alignment with other members of the MFS and construction of a neighbour-joining tree. This strategy has been successfully used to predict the substrate and mechanism of transporters (Martin *et al.* (2005)) (Saliba *et al.* (2006)).

Seventeen BLAST iterations were performed using PAD1 as a query sequence until the results came to convergence and no more hits were returned. In the top 250 hits, 19 proteins had putative or characterised transport substrates identified (see Table 3.8-1). Of these, 15 were oxalate/formate antiporters, 2 were non-specific

monocarboxylate transporters, 1 was a nitrate and chloride transporter and 1 was a DNA polymerase III alpha subunit. Oxalate, $(COO)_2^{2}$, and formate, $HCOO^-$, are di-carboxylates and monocarboxylates respectively. Nitrate, NO_3^-, is a small inorganic anion with a formal charge of -1. Note that the return of the alpha subunit of DNA polymerase III is probably a miss-annotation of the database; a Pfam analysis shows it to be a member of the MFS and it has no polymerase domains. Moreover, a TBLASTX search using this gene as a query sequence returned only transporters (data not shown).

The return of predominantly carboxylate transporters may be significant because both citrate and *cis*-aconitate are tricarboxylates, and a role for the PAD proteins in citrate or *cis*-aconitate transport correlates with a role in citrate /*cis*-aconitate mediate differentiation. The return of 15 oxalate/formate antiporters may also indicate that PAD1 also functions as an antiporter.

Table 3.8-1 A summary of the characterised hits from a PSI-BLAST using PAD1 as a query sequence.

Accession Number	Description	E value
gi|57545995|gb|AAW51884.1|	nitrate and chloride transporter [Glycine max]	8E-130
gi|152982869|ref|YP_001355198.1|	oxalate/formate antiporter, MFS superfamily [Janthinobacterium sp. Marseille].	1.00E-95
gi|116695641|ref|YP_841217.1|	Oxalate / formate antiporter, MFS-type, OFA family [Ralstonia eutropha H16]	4.00E-92
gi|39935013|ref|NP_947289.1|	possible oxalate/formate antiporter [Rhodopseudomonas palustris CGA009]	1.00E-89
gi|148254538|ref|YP_001239123.1|	putative oxalate/formate antiporter [Bradyrhizobium sp. BTAi1]	1.00E-89
gi|146341834|ref|YP_001206882.1|	putative oxalate/formate antiporter [Bradyrhizobium sp. ORS278]	7.00E-89
gi|152982521|ref|YP_001354332.1|	oxalate/formate antiporter, MFS superfamily [Janthinobacterium sp. Marseille]	2.00E-86
gi|152982240|ref|YP_001353146.1|	oxalate/formate antiporter, MFS superfamily [Janthinobacterium sp. Marseille]	4.00E-86
gi|146341835|ref|YP_001206883.1|	putative oxalate/formate antiporter [Bradyrhizobium sp. ORS278]	6.00E-86
gi|27378260|ref|NP_769789.1|	putative Oxalate:formate antiporter [Bradyrhizobium japonicum USDA 110]	5.00E-84
gi|90020642|ref|YP_526469.1|	DNA polymerase III, alpha subunit [Saccharophagus degradans 2-40]	7.00E-84
gi|148254537|ref|YP_001239122.1|	putative oxalate/formate antiporter [Bradyrhizobium sp. BTAi1]	7.00E-84
gi|114327455|ref|YP_744612.1|	oxalate/formate antiporter [Granulibacter bethesdensis CGDNIH1]	1.00E-83
gi|152982449|ref|YP_001355163.1|	oxalate/formate antiporter, MFS superfamily [Janthinobacterium sp. Marseille]	4.00E-81
gi|152980784|ref|YP_001354232.1|	oxalate/formate antiporter, MFS superfamily [Janthinobacterium sp. Marseille]	5.00E-75
gi|134095635|ref|YP_001100710.1|	putative oxalate/formate antiporter [Herminiimonas arsenicoxydans]	2.00E-74
gi|70997153|ref|XP_753331.1|	MFS monocarboxylic acid transporter, putative [Aspergillus fumigatus Af293]	1.00E-62
gi|121713978|ref|XP_001274600.1|	MFS monocarboxylic acid transporter, putative [Aspergillus clavatus NRRL 1]	1.00E-61
gi|91781396|ref|YP_556602.1|	Putative oxalate/formate antiporter, majorfacilitator superfamily (MFS) [Burkholderia xenovorans LB400]	2.00E-55

Chapter 4

Characterisation of the DiD1 cell line

4 Characterisation of the DiD1 cell line

As described in Section 1.8, the DiD1 differentiation phenotype was selected for by selecting for cells that failed to lose their protective VSG coat in response to *cis*-aconitate (Tasker *et al*. (2000)). Deletion of GPI-PLC caused a reduced virulence, and an unexpected reversion to pleomorphism, consequently restoring DiD1 cells' ability to respond to *cis*-aconitate. The underlying defect in the DiD1 cells is not known, but it was proposed that they lack a stumpy form characteristic enabling them to differentiate to the procyclic form. Thus, an analysis of the defect in the DiD1 cell line may give insight into the mechanisms that allow the initiation of differentiation.

4.1 A Northern analysis to compare PAD genes in the DiD1 and parental cell lines

As described in Section 1.8, a macroarray hybridisation analysis identified a plasmid containing fragments of PAD1 and 2 as differentially hybridising to DiD1 cDNA with respect to the parental cDNA. Subsequent Northern analysis using a riboprobe based upon this plasmid verified the original result (Davies and Levin, University of Manchester, unpublished data). As discussed in Section 3.1, the PAD array constitutes 8 highly homologous genes at the 3' end of a directional gene cluster. Thus, it was of interest to determine if PAD1, PAD2 or both genes were enriched in the DiD1 cell line, and if other members of the array were also differentially expressed.

Riboprobes that were synthesised previously (courtesy of Pam Davies, University of Manchester) were used in a Northern blot to determine if there was differential expression between the parental and the DiD1 cell line. These probes were designed to detect:

- PAD1

- PAD2
- PAD3
- PAD5 and 7 (the nucleotide sequence for these genes is identical, thus it is not possible to distinguish between them)

Additionally, a riboprobe was designed to detect all 8 genes simultaneously by using the entire open reading frame of PAD2.

Figure 4.1-1 shows that, of the mRNA transcripts tested, only PAD1 and PAD2 were found to be enriched in the DiD1 cell line; PAD3 mRNA was not detected (data not shown), suggesting that this transcript is either not expressed or that the riboprobe preparation failed (no further analysis was performed on PAD3); PAD4, 6 and 8 expression was not assayed. The riboprobe to detect the entire array of PAD also detected an upregulation of PAD genes in DiD1, but this was mild in comparison with PAD1 and PAD2 upregulation, suggesting that few of the PAD genes had enriched expression in the DiD1 cell line.

Thus, it appears that PAD1 and PAD2 are probably the only members of the PAD array that are aberrantly expressed in the DiD1 cell line. This may not be surprising given that these genes are adjacent within the array. The 3' UTRs of PAD1 and PAD2 are distinct (see Section 3.3.2), suggesting that the upregulation of these transcripts may not be due to a single mRNA binding protein that confers stability upon both transcripts. Moreover, despite the 3' UTRs of PAD5 and 7 sharing a 272 bp element with PAD1 (see Section 3.3.3), PAD5 and 7 were not enriched in the DiD1 cell line, implying that enrichment of PAD1 in DiD1 cells was not conferred by this region of the 3' UTR.

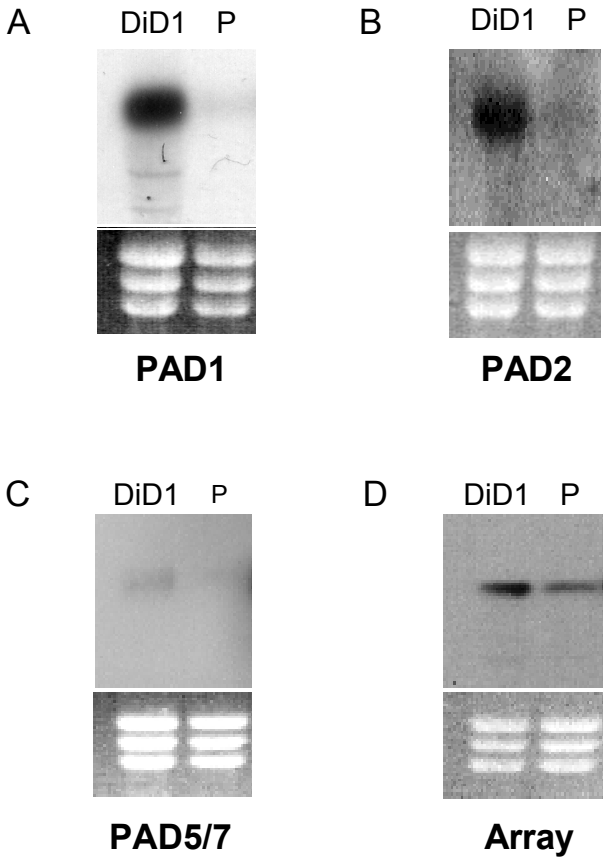


Figure 4.1-1 Expression of PAD genes in parental and DiD1 cells. A Northern blot analysis to determine expression level differences between the parental (P) and DiD1 cell line of (A) PAD1, (B) PAD2, (C) PAD5/7 and (D) the entire array, shows that only PAD1 and PAD2 express elevated levels of mRNA transcript in DiD1 cells.

4.2 A Southern analysis to detect PAD array genome rearrangements

PAD1 and PAD2 may have been subject to a genomic re-arrangement that has caused them to be duplicated and consequently more highly expressed, or they may have moved to a region of the genome that is more highly transcribed. To test these possibilities, a Southern blot was performed to detect any large scale genomic re-arrangements affecting the PAD array genes in the DiD1 line.

The original DiD1 cell line was culture adapted subsequent to its selection in order to generate GPI-PLC deletion mutants. However, this cell line was lost in a freezer malfunction and subsequent attempts to culture-adapt the DiD1 cell line were unsuccessful. For this reason, DiD1 and the parental cells were purified from infected mouse blood using a DEAE cellulose anion exchange column, and genomic DNA was prepared from the cell lysate (see Materials and Methods for details). Parental and DiD1 genomic DNA were subsequently digested with either EcoRI or BamHI for Southern blotting.

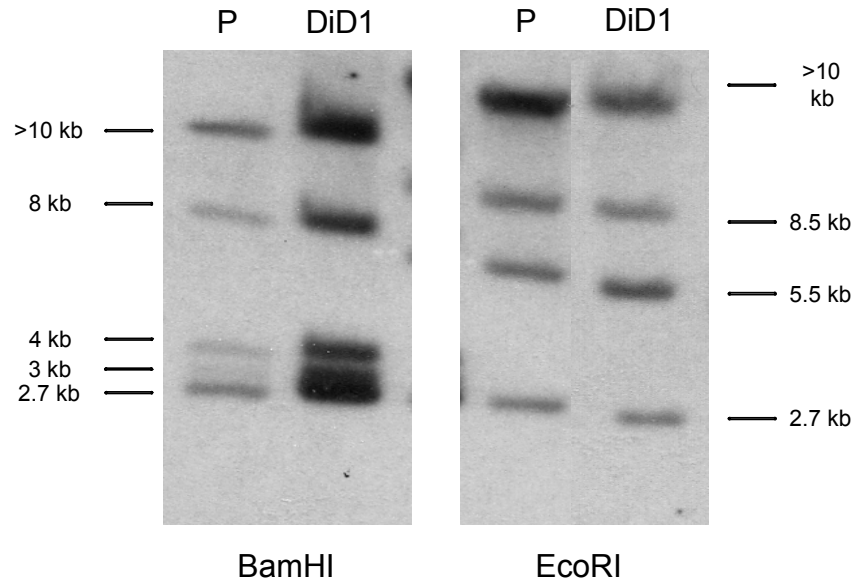
A DNA probe was synthesised by treating an amplicon containing the entire open reading frame of PAD2 with Klenow polymerase in the presence of labelled nucleotides (see Materials and Methods for details). The homology of the PAD genes is such that this probe would anneal to all of the PAD genes (see Section 3.2 for PAD gene alignments).

Figure 4.2-1 shows that, when this probe was used in a Southern blot, there did not appear to be any differences in probe hybridisation between the DiD1 cell line and the parental strain; however, the bands that were observed do not correlate precisely with the sizes predicted based on the genome reference data. The database predicts that 9 fragments should be released through BamHI digestion; of these, it should be possible to distinguish at least 6 on a 0.7% agarose gel, but only 5 bands were detected. Additionally, when the genomic DNA is digested using EcoRI, 3 bands are predicted from the database but 4 are detected. It is likely that differences between the genome reference strain and the parental cell line have accumulated since the

strains diverged, resulting in modifications of restriction enzyme sites, or additions and deletions of genomic DNA sequence. This is not uncommon in arrays of tandem repeats; in their analysis of 3 different *T. brucei* strains (427, 247 and the 927 genome reference strain), Callejas *et al*. (Callejas *et al.* (2006)) report that either duplications or deletions had occurred in all of the tandem gene arrays in chromosome 1. Of the bands that were detected, the sizes of the predicted bands appeared to correlate approximately with those that were observed, suggesting that differences are minor. Nonetheless, it is clear that there are differences between the genome reference strain and the EATRO 2340 GUP2965 cell line in this region of the chromosome.

There were, however, no differences detected between the DiD1 cell line and the parental cell line. We can therefore conclude that it is unlikely that large scale genomic rearrangements have occurred in this region of the genome during the selection of DiD1 cells. It should be noted, however, that if the entire array has been duplicated along with the restriction sites used in this analysis, this may not be detected by Southern blot.

A



B

BamHI: Exp.	Obs . - P	Obs . - DiD1
11.1	+10	+10
10.1	+10	+10
6.3	8	8
3.3	4	4
3.2	3	3
3.2	2.7	2.7
3.0		
3.0		
2.7		

EcoRI: Exp.	Obs . - P	Obs . - DiD1
23.6	+10	+10
8.7	8.5	8.5
5.6	5.5	5.5
	2.7	2.7

Figure 4.2-1 A Southern blot to determine whether any genomic rearrangements have occurred during the selection of the DiD1 cell line. (A) A Southern blot analysis performed on parental (P) and DiD1 genomic DNA shows that a large scale genomic rearrangement has not occurred during DiD1 selection. (B) A summary of the Southern data, compared with what would be predicted genome reference data. There does appear to be some differences between the observed and the expected bands.

4.3 A sequencing analysis to compare the PAD1 3' intergenic region in the DiD1 and parental cell lines

Trypanosomatids control mRNA transcript levels predominantly through the use of mRNA binding proteins that bind to the 3' UTR of expressed genes. These predicted proteins can confer stability or instability (Schürch *et al.* (1997)) (Furger *et al.* (1997)) (Hotz *et al.* (1997)) (Mayho *et al.* (2006)), as well as either enhancing or repressing translation of the target transcript. As discussed (see Section 4.1), PAD1 and PAD2 mRNA transcripts were found to be enriched in the DiD1 cell line, and a Southern analysis concluded that this was not due to an obvious genomic rearrangement (see Figure 4.2-1). Therefore, an alternative possibility was that enrichment of PAD1 and 2 expression in DiD1 cells may be due to the presence of mutation(s) in the 3' UTR of these genes during selection of the DiD1 cell line, thus preventing regulation of the mature transcript. A sequence analysis was therefore performed on the PAD1 3' UTR. The PAD1 3' UTR was chosen rather than PAD2 3' UTR because the data from the Northern analysis indicated that this gene has a role in differentiation (being both upregulated in DiD1 cells (see Section 4.1) and stumpy form cells (see Section 5.1)), and, as it has the smallest 3' UTR, it was relatively straightforward analyse.

Primers were designed that were predicted to amplify the entire intergenic sequence between PAD1 and PAD2 (see Appendix A for primer list). Genomic DNA from the parental cell line and from the DiD1 cell line was used as a template, and the resulting amplicons were cloned into pGEM-T easy for DNA sequencing (see Materials and Methods for details). To prevent the possible mis-interpretation of a polymerase-incorporated error, two separate reactions were performed for each template, and 5 positive clones derived from each PCR reaction were sequenced to make a total of 10 clones for each template. A consensus sequence of each clone was then aligned with the genome reference data.

Figure 4.3-1 shows that no mutations in the PAD1 – 2 intergenic region were detected in the DiD1 cell line. From ten DiD1 clones, one clone had a T → A

change at position 140, and one clone had a G → T change at position 178 (data not shown); these were considered to be polymerase-incorporated errors and were not detected in any of the other clones. Additionally, from ten parental clones, one clone had a T → C difference at position 789, and one clone had a T → G difference at position 821; similarly, these were considered to be polymerase incorporated errors as they were not present in any of the other clones (data not shown). Interestingly, this analysis did determine that there are 2 'alleles' of this intergenic region (discriminated by G/A at position 929) in both the parental strain and the DiD1 cell line. The genome reference strain appears to represent a third 'allele'.

Thus it would appear that the enrichment of PAD1 is not due the presence of mutation(s) within the PAD1 3' UTR. A similar analysis of the PAD2 3' UTR was not thought to be necessary as one would expect the enrichment of these genes in the DiD1 cell line to have a common cause. The underlying cause of PAD1 and PAD2 mRNA transcript enrichment in the DiD1 cell line is therefore not known. One possibility is that altered specificity or action of a PAD gene RNA binding protein has occurred through the selection process that prevents proper de-stabilisation of these transcripts in the monomorphic slender form. If this is the case, then one could speculate that, despite the distinct expression profile of these genes, both are acted upon by a common RNA binding protein, possibly as part of a multi-protein complex. It would be informative to determine which RNA binding proteins regulate the different PAD genes, and to determine whether there is any altered RNA target specificity in the DiD1 cell line.

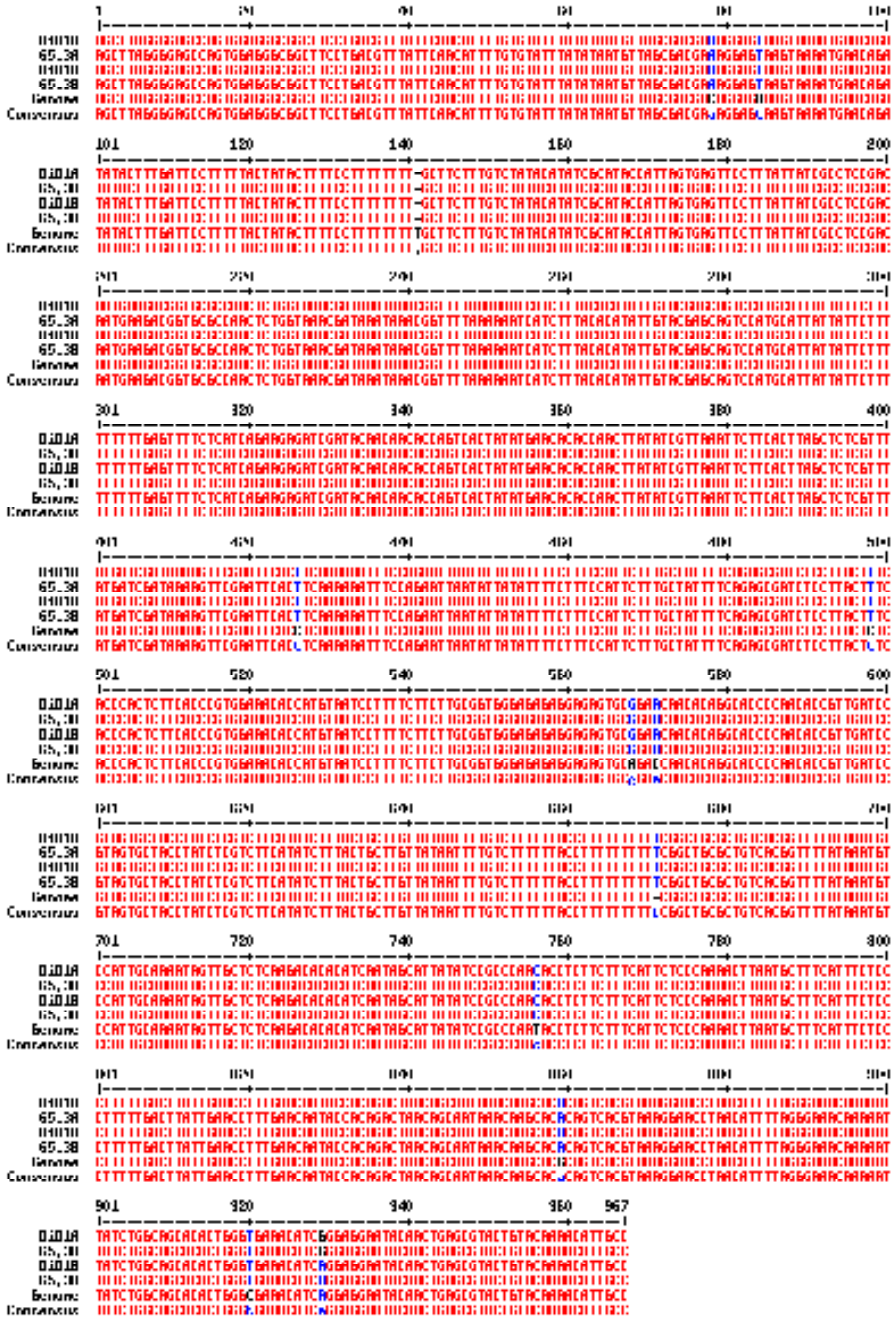


Figure 4.3-1 A sequencing analysis of the PAD1 3' intergenic region. The PAD1 3' intergenic region was amplified from the DiD1 and 65.3 parental strain and cloned for sequencing. 10 clones from 2 different PCR reactions were sequenced to ensure that both copies of the PAD1 3' UTR were analysed, The resulting 'alleles' were aligned with the PAD1 3' intergenic region from the genome reference data.

4.4 A sequencing analysis to determine the presence of mutations of PAD genes in the DiD1 cell line

The enriched expression of PAD1 mRNA in stumpy form cells (see Section 5.1) suggests that this gene may have a role in initiating differentiation to the procyclic form. Additionally, PAD2 protein is elevated in expression in response to conditions that induce hypersensitivity to *cis*-aconitate (see Section 5.16), supporting its putative role in promoting *cis*-aconitate mediated differentiation. It is thus counter-intuitive that PAD1 and PAD2 have elevated expression in a cell line that is defective in its ability to respond to *cis*-aconitate. One possibility is that the selection regime favoured cells with defective PAD proteins that cause a dominant negative effect. If this were the case, one would expect that PAD genes in DiD1 have accumulated mutation(s) through the selection cycles. To test this hypothesis, a sequence analysis was performed to detect any differences in PAD1 and PAD2 between the DiD1 and the parental cell lines.

The polymerase chain reaction (PCR) was used to amplify PAD1 and 2 from DiD1 and parental genomic DNA (see appendix A for primer list, and Materials and Methods for more details). As with the cloning of the intergenic regions (see Section 4.3), two separate reactions were performed to eliminate the possibility of a polymerase incorporated confusing the analysis. The amplicons corresponding to PAD1 and 2 were cloned into pGEM T easy and 5 positive clones from each PCR reaction (in total 10 from 65.3, and 10 from the DiD1 cell line, *per* gene) were submitted for sequencing.

Figure 4.4-1 shows that that there are 2 PAD1 alleles in the 65.3 parental strain, distinguished by 11 amino acid polymorphisms. Additionally, the genome reference strain appears to represent a third allele. Both alleles were found in the DiD1 cell line, and no mutations appear to have occurred during the DiD1 selection.

Figure 4.4-2 shows that, as with PAD1, there are 2 PAD2 alleles in the parental cell line. These can be distinguished by 2 amino acid polymorphisms; again, a distinct

allele is present in the genome reference strain. As with PAD1, both alleles were detected in the DiD1 cell line, and neither appeared to have accumulated any mutations.

In summary, no mutations were detected in the DiD1 PAD1 or PAD2 alleles that might cause a dominant negative effect. Given that no mutations were detected by this analysis in the 3' UTR of PAD1, and Southern blot did not reveal any genomic re-arrangements in the PAD array locus, upregulation of PAD1 and 2 in the DiD1 cell line is probably a consequence of an unknown defect that caused the DiD1 cells to differentiate poorly in response to *cis*-aconitate.

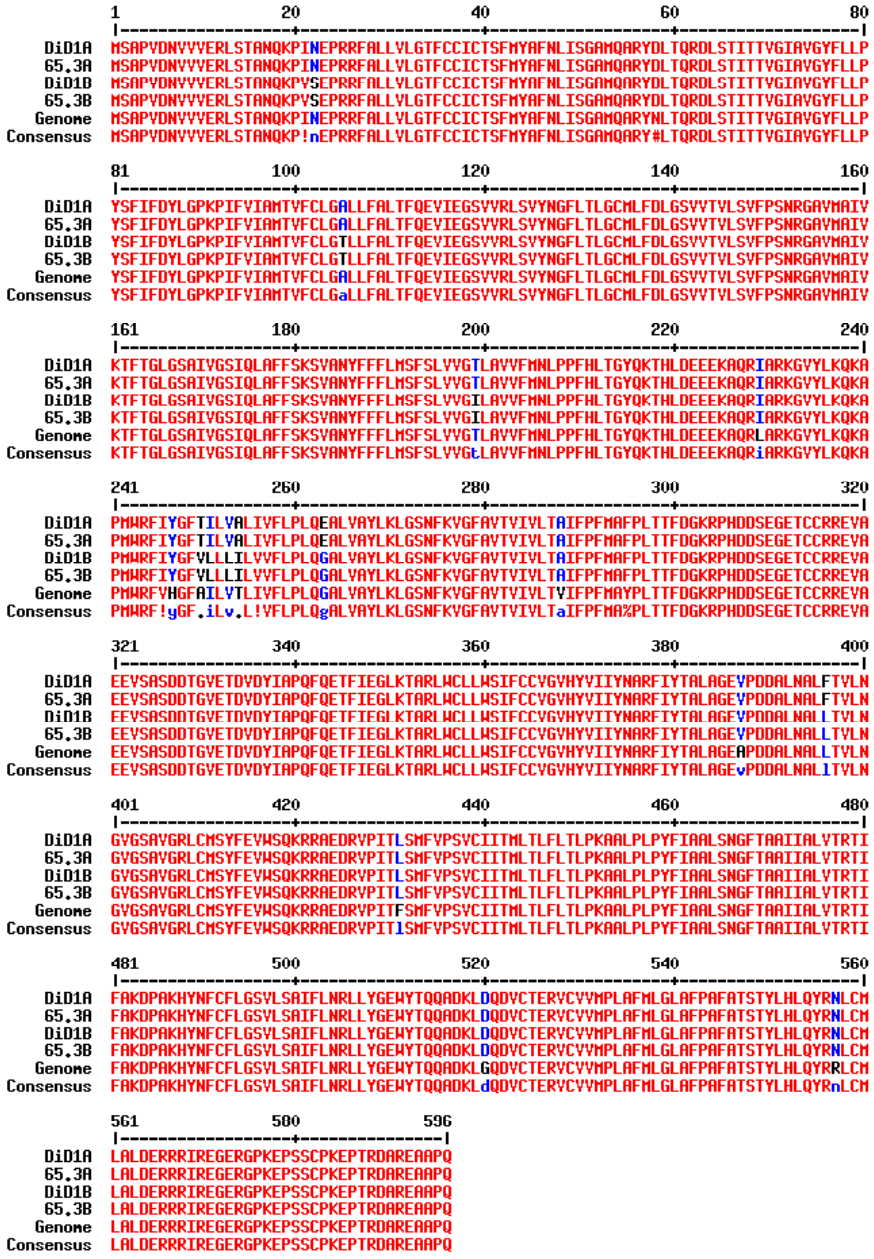


Figure 4.4-1 An alignment of the parental, DiD1 and genome reference strain PAD1 encoded proteins. PAD1 was amplified from DiD1 and parental 65.3 genomic DNA and cloned into pGEM T easy for sequencing. Consensus protein sequences were generated and aligned with the PAD1 genome reference sequences.

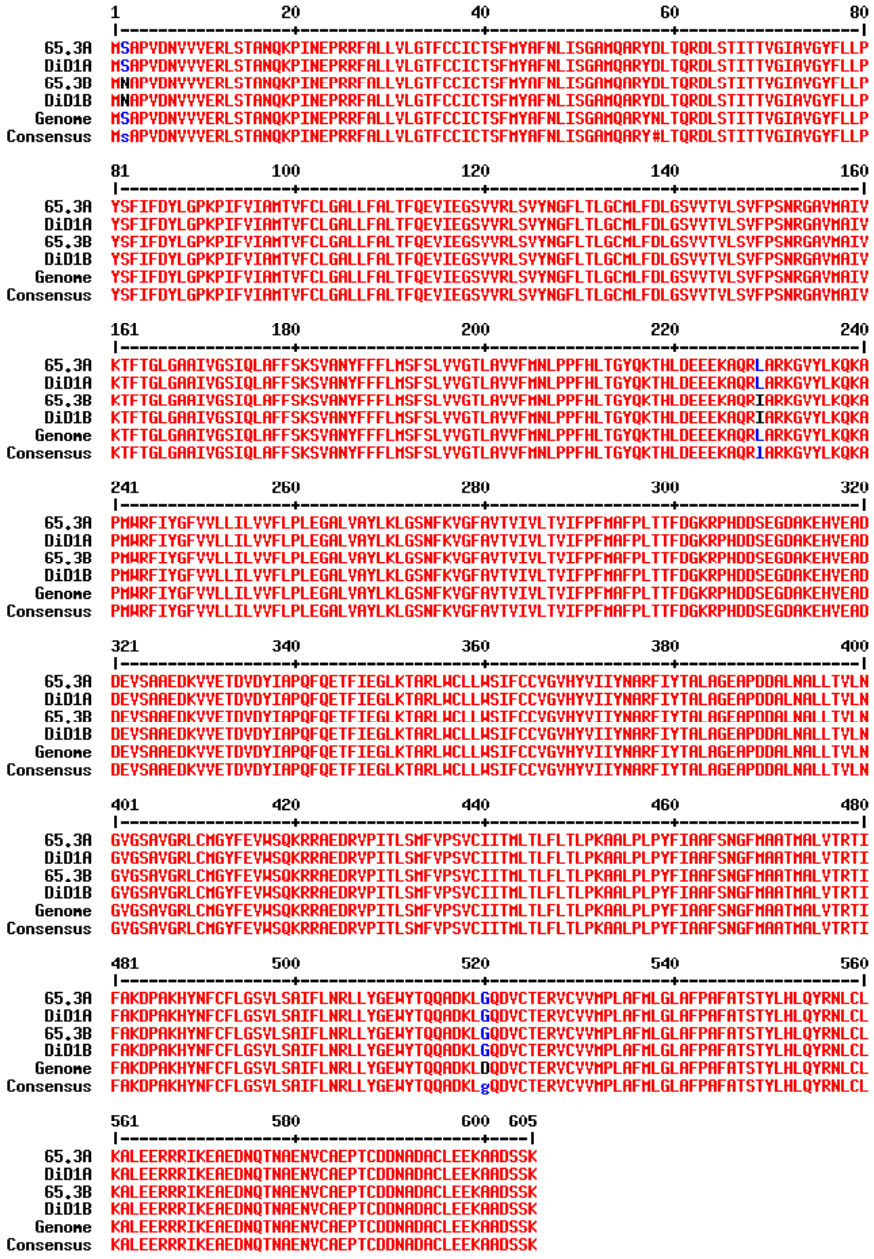


Figure 4.4-2 An alignment of the parental, DiD1 and genome reference strain PAD2 encoded proteins. PAD2 was amplified from DiD1 and parental 65.3 genomic DNA and cloned into pGEM T easy for sequencing. Consensus protein sequences were generated and aligned with the PAD2 genome reference sequences

4.5 Ectopic expression of PAD1 to recapitulate the DiD1 differentiation defect

The mechanism of PAD1 and PAD2 upregulation in DiD1 cells was not determined by any of the analyses presented in this study. Additionally, the sequence analysis of PAD1 and PAD2 had not uncovered any evidence of a mutated PAD gene that might cause a dominant negative effect. It was therefore of interest to determine whether upregulation of PAD genes *per se* was sufficient to inhibit differentiation of cells. A possible, speculative, mechanism for this might be that low levels of PAD protein are able to transmit the *cis*-aconitate signal to differentiate in monomorphic slender cells, but high mRNA expression causes mal-folding of PAD proteins such that no functional protein is formed, thus inhibiting differentiation.

The following experiment set out to test this possibility.

4.5.1 Over expression of PAD1

PAD1 was amplified from parental 65.3 genomic DNA and cloned into the expression vector pHD451-Ty. This DNA construct had been previously modified from pHD451 such that insertion of a gene with no stop codon into the multiple cloning site incorporates a carboxyl terminal Ty epitope tag upon expression (Bastin *et al*. (1996)) (Hendriks *et al*. (2001)). This construct was further modified to facilitate cloning (see Section 5.2.1.1). Positive clones were screened by diagnostic digest and verified by DNA sequencing (see Materials and Methods). Monomorphic slender *T. brucei* Lister 427 cells containing the pHD449 TetR constitutive expression construct (Biebinger *et al*. (1997)) were then transfected with the pHD451:PAD1-Ty (see Materials and Methods for details) such that cells would express PAD1-Ty in the presence of tetracycline.

The differentiation of cells that had been induced to express PAD1-Ty was compared with those that had not. Additionally, a non-transfected cell line (with and without tetracycline) was included in case there was low level ('leaky') expression of PAD1-Ty even in non-induced cells, or an effect of tetracycline unrelated to PAD1-Ty

expression. Cells were induced for PAD1-Ty expression by supplementing the culture media with 2 μg ml^{-1} tetracycline for 3 days, and differentiated by incubating at 27°C in HMI9 supplemented with 6 mM *cis*-aconitate. Differentiation capacity was assayed by fixing cells in methanol at 0, 24, 48 and 72 hours post *cis*-aconitate, and staining for the presence of EP procyclin on the cell surface membrane (see Materials and Methods for details). The level of PAD1-Ty expression was determined by Western blot analysis of protein cell lysate made just prior to the addition of *cis*-aconitate using the anti-TY BB2 antibody.

Figure 4.5-1A shows that inducing the expression of PAD1-TY does not have an effect upon differentiation, thus no dominant negative, or enhancement, effect was observed. The Western blot (Figure 4.5-1B) shows that PAD1-Ty was being expressed in the presence of tetracycline, although the corresponding band is faint. This weak signal could be due to a number of reasons; transporter proteins have a complex structure, and aberrantly high expression may cause mis-folding, and consequent degradation. Additionally, the tight packing of VSG on the cell surface may mean that high expression of a cell surface membrane bound protein was not possible. Alternatively, the weak signal could be due to technical difficulties associated with detection of hydrophobic proteins.

PAD1 mRNA expression levels between the DiD1 cell line and the expression cell line were not tested, but a comparison may determine whether the expression levels are equivalent, and thus whether the inability to re-capitulate the DiD1 phenotype is due to low expression of PAD1, or whether the enrichment of PAD1 transcript in the DiD1 cell line is a symptom of a different defect. One possibility is that the Ty epitope tag interfered with PAD1 function, thus preventing a dominant negative effect from occurring; later experiments made use of ectopic expression of an untagged gene, and also titrated the levels of *cis*-aconitate to determine whether an effect of PAD1 expression was detected at low *cis*-aconitate concentrations (see Section 6.1).

Nonetheless, it was clear that inducing the expression of PAD1-Ty in monomorphic cells did not detectably enhance, or repress, differentiation.

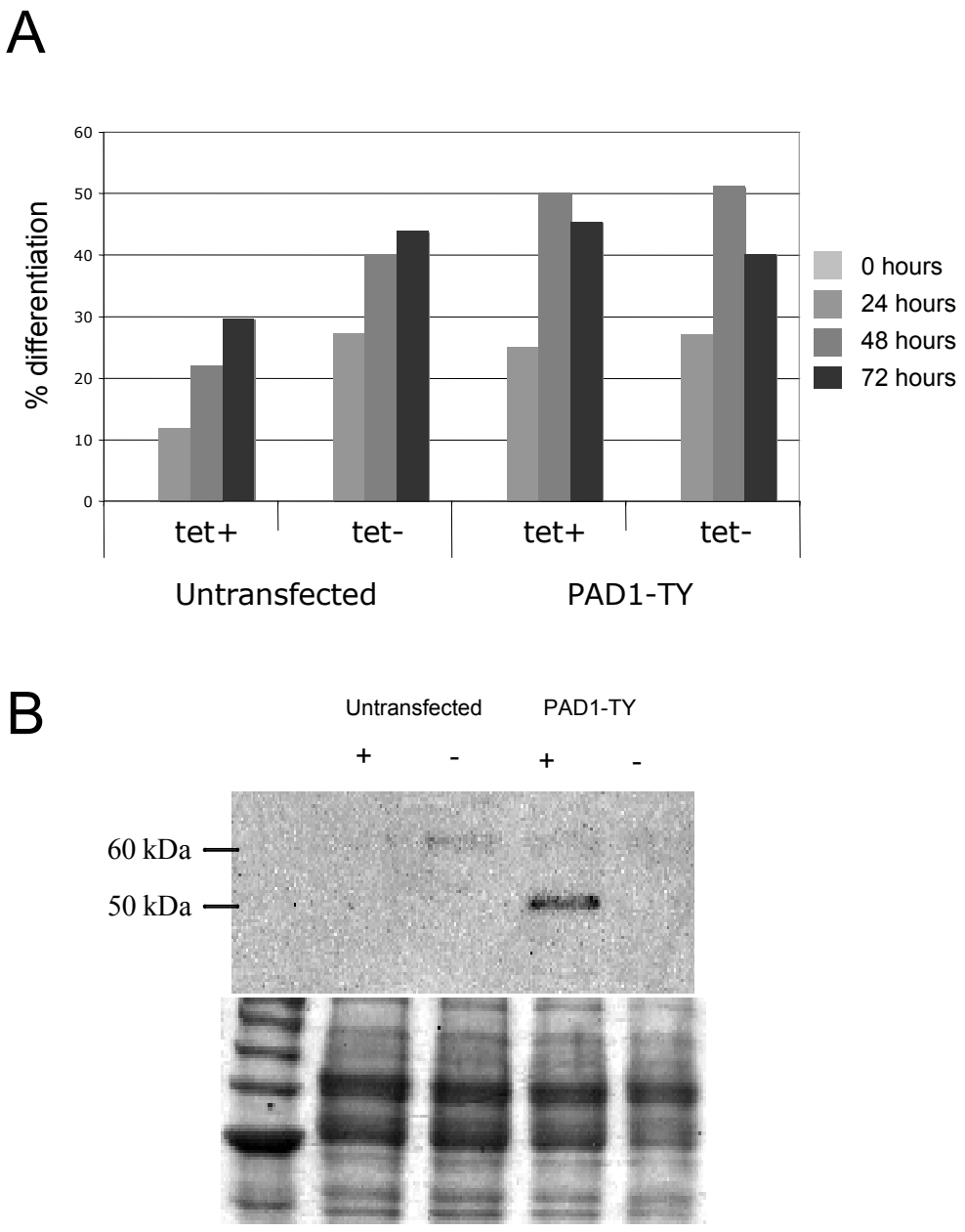


Figure 4.5-1 The effect of PAD1 upon differentiation in response to *cis*-aconitate. Monomorphic slender cells ectopically expressing an epitope tagged PAD1 in response to tetracycline gene were assayed for their ability to differentiate. (A) Untransfected cells and cells containing an ectopic copy of PAD1 fused to a TY tag were assayed for their ability to differentiate in the presence and absence of tetracycline. (B) A Western blot using the BB2 monoclonal antibody demonstrates that transgenic cells express the Ty epitope tagged PAD1 in response to tetracycline.

4.6 A Western to determine whether PAD proteins are upregulated in the DiD1 cell line

As mRNA expression is not always reflected in protein levels, it was of interest to determine whether PAD1 and 2 protein are also being expressed in DiD1 cells in addition to their mRNA. If, despite the high levels of mRNA transcript, PAD1 and PAD2 proteins were not detected, then one might speculate that the differentiation defect was due to inability to synthesise these proteins.

DiD1 cell lysate was analysed for endogenous PAD array expression by Western blot using the α-PAD1 #1 antibody to detect PAD1, and the α-array #2 antibody to detect all of the PAD proteins at once. The design and verification of these antibodies is discussed in depth in Section 5.2.2.1 and Section 5.5.1; briefly, peptides containing an amino acid sequence common to all of the PAD proteins, and peptides containing an amino acid sequence specific to PAD1, were used to immunise rabbits, and the resulting PAD specific antibodies were purified using the corresponding immunogenic peptides. The LI-COR Odyssey system was used for quantitation relative to a tubulin loading control (see Materials and Methods for details). Parental cell protein was included for comparison, and stumpy form protein was included as a positive control.

Figure 4.6-1A shows that PAD1 was expressed in DiD1 cells and not in the parental strain, consistent with the Northern blot data (see Section 4.1). However, quantitation demonstrated that Stumpy form cells expressed PAD1 approximately 20 fold more than DiD1 cells.

Using the α-array antibody, Figure 4.6-1B shows that 2 PAD proteins were expressed in DiD1 cells, whereas no PAD proteins were detected in the parental 65.3 cell line. The 55 kDa band was subsequently shown to correspond to PAD1, whereas the 57 kDa band was shown to correspond to PAD2 (see Section 5.5.3).

Thus, it would appear that the Western blot data is consistent with the Northern data, and that both PAD1 and PAD2 are expressed in DiD1 cells. We can therefore conclude that the DiD1 defect is not due to the inability of DiD1 cells to express PAD proteins. This result further highlights the difference between the DiD1 cells, and non-selected monomorphic parental cells.

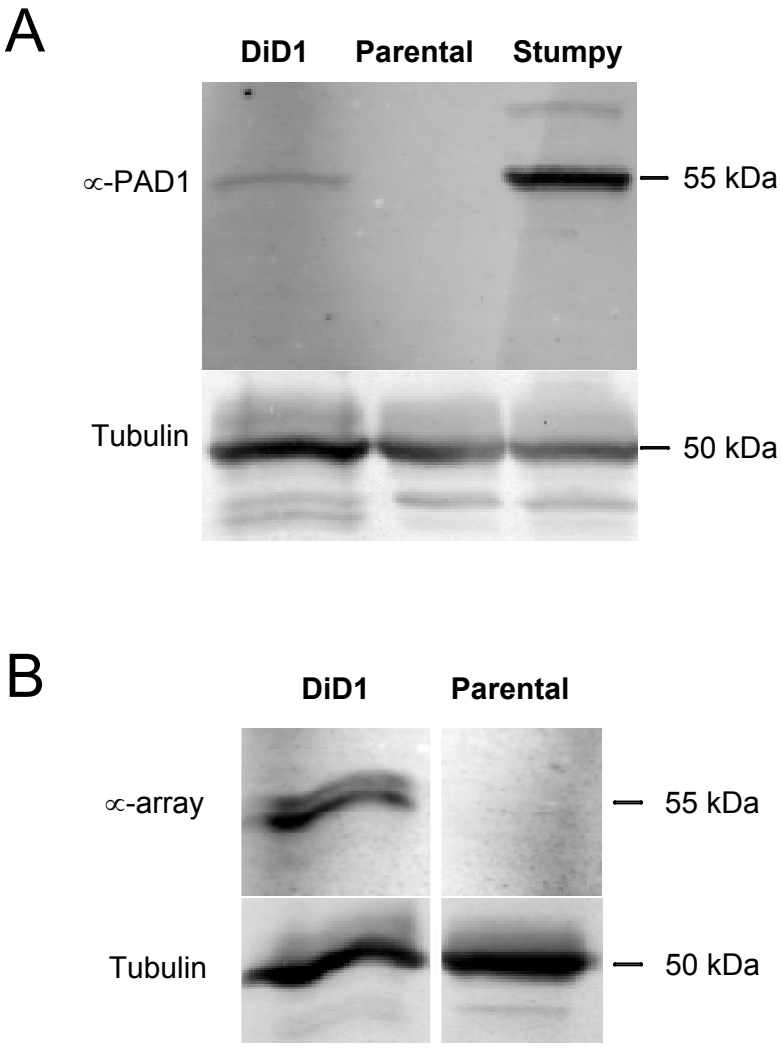


Figure 4.6-1 PAD proteins are expressed in DiD1 cells. (A) DiD1, parental 65.3 and stumpy form protein was analysed by Western blot with α-PAD1 to determine whether PAD1 is expressed in DiD1 cells. PAD1 is not expressed in the parental cells but is expressed in DiD1 cells. Expression is 20 fold higher in stumpy form cells than DiD1 cells (B) Western blot of DiD1 and parental 65.3 protein showed that 2 PAD proteins were expressed in DiD1 cells, and that none was detected in the parental cells. Note that in (B), the protein was loaded on non-adjacent lanes, and that the intervening lanes are not shown. The exposure, however, is equivalent.

4.7 A Northern analysis to determine whether there is a global upregulation of stumpy-enriched transcripts in the DiD1 cell line

The original study of the DiD1 cells (Tasker *et al.* (2000)) showed that these cells are monomorphically slender and do not highly express the enzyme DHLADH, an enzyme known to be enriched in stumpy form cells due to their mitochondrial activation. However, other stumpy form enriched proteins were not examined. If a global upregulation of stumpy form enriched proteins was found in the DiD1 cell line, this would provide insight into the mechanism underlying the DiD1 defect. The macroarray hybridization analysis to assay for upregulation of transcripts was based upon a sheared genomic library and thought to have approximately 80% genome coverage (Elisabetta Ullu, University of Yale, personal communication); thus many genes may have been excluded from this analysis.

4.7.1 TbPTP1

The protein tyrosine phosphatase TbPTP1 was shown to be enriched in stumpy form cells (Szöor *et al*. (2006)). This protein is thought to act as a 'break' upon differentiation, and when inactivated by the tyrosine phosphatase specific inhibitor 3-(3,5-dibromo-4-hydroxy-benzoyl)-2-ethyl-benzofuran-6-sulfonicacid-(4-(thiazol-2-ylsulfamyl)-phenyl)-amide (referred to as BZ3) blood form cells spontaneously differentiate. Therefore one can speculate that any upregulation of a group of proteins including TbPTP1 would inhibit the ability of these cells to initiate differentiation. A Northern blot using a DIG labeled riboprobe specific to TbPTP1 was therefore performed on slender, stumpy, procyclic and DiD1 total mRNA (see Materials and Methods for details).

Figure 4.7-1A shows that there was no convincing upregulation of TbPTP1 mRNA transcript detected in DiD1 cells This demonstrated that the DiD1 defect was not a result of an upregulated PTP1 applying a constitutive break upon differentiation. This does not exclude the possibility that a mutant PTP1 is constitutively active in DiD1 cells.

4.7.2 ESAG-K9 and glutamine synthase

ESAG-K9 and glutamine synthase were found to be enriched in stumpy form cells (Keith Matthews, unpublished data). Therefore, their relative expression level in the parental 65.3 and DiD1 cells was determined to assay whether there is a global upregulation of stumpy form enriched transcripts in the DiD1 cell line.

A Northern blot using DIG labelled riboprobes on 65.3 parental and DiD1 total RNA was performed to determine whether ESAG-K9 and glutamine synthase transcripts were enriched in DiD1 cells. Stumpy form total RNA was included for comparison. Figure 4.7-1B shows that no significant upregulation of either of these transcripts was detected.

Thus, a Northern blot analysis to detect expression of PTP1, ESAG-K9 and glutamine synthase in parental cells versus DiD1 cells showed that there was no detectable enrichment of these transcripts in the DiD1 cells. Moreover, Figure 4.1-1 demonstrated that PAD5 and 7, both highly enriched in stumpy form cells (see Section 5.1), were not appreciably expressed in DiD1 cells. Therefore, there does not appear to be a global upregulation of stumpy enriched transcripts. The cause of the DiD1 defect therefore remains unidentified.

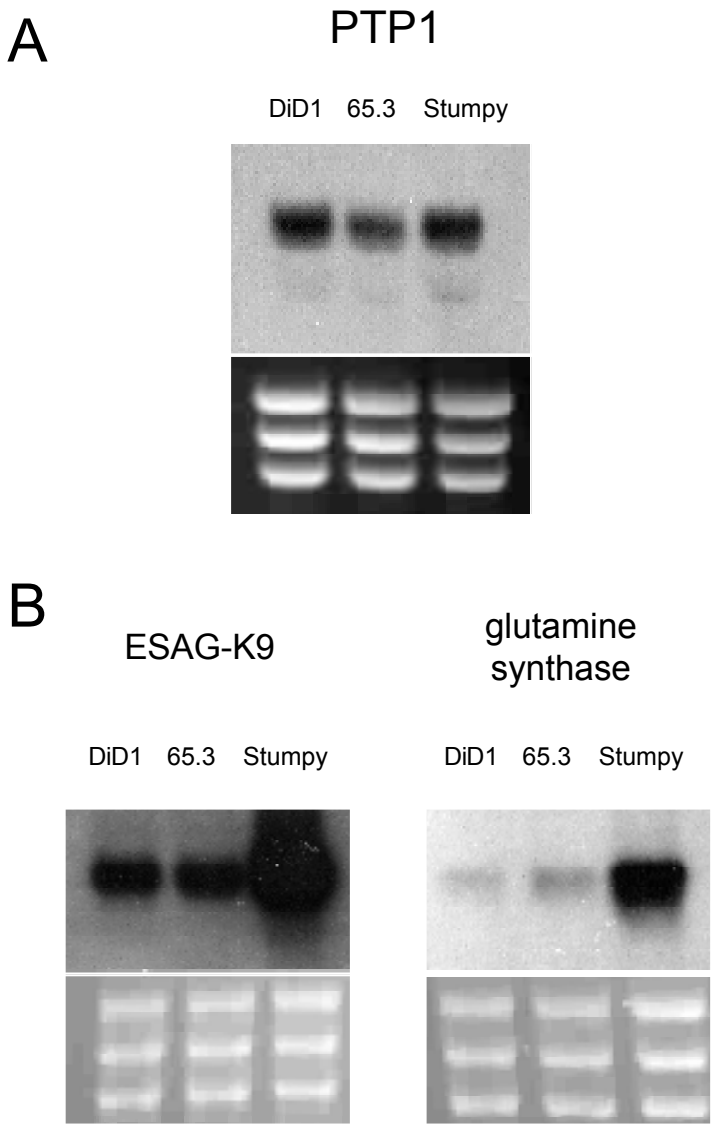


Figure 4.7-1 There is not a global stumpy form-transcript upregulation. A Northern blot was performed on DiD1, parental 65.3 and stumpy form RNA to determine whether there was an upregulation of (A) PTP1, and (B) ESAG-K9 and glutamine synthase. No significant upregulation was detected.

Chapter 5

Expression and Localisation of the PAD genes

5 A molecular characterisation of the PAD array

The tools for molecular analyses of *Trypanosoma brucei* are well developed and sophisticated. The genetic pliability of *T. brucei* has resulted in the routine use of RNAi mediated transcript ablation, ectopic expression of *trans* genes, and generation of deletion mutants by homologous recombination. The completion of the *T. brucei* genome project has provided the trypanosome community with a wealth of data, as well as a valuable tool for bioinformatics analysis.

However, forward genetics has, of necessity, been limited because the epimastigote form that is thought to undergo genetic recombination is not amenable to culture, and trypanosomes do not easily complete their lifecycle in tsetse. Thus, identifying interesting and important genes beyond the analogues of characterised proteins is a major challenge in trypanosome biology. The selection for the DiD1 phenotype, and the subsequent identification of the PAD array, represents an innovative use of a forward genetics approach to identification of genes with a putative function.

With the identification of the PAD genes, and their implied function in differentiation, the molecular toolkit described above was employed to dissect their role in *T. brucei* lifecycle progression.

5.1 A Northern analysis to determine the RNA expression profile of the PAD genes

The analysis of the PAD array genomic region showed that, although the coding sequences of the PAD genes are highly homologous, the intergenic regions and the predicted 3' UTRs that likely control mRNA transcript stability have significant variability. In particular, the PAD2 predicted 3' UTR is unique within the PAD array, whereas PAD1, 5 and 7 have similar repetitive elements (see Section 3.3.3). PAD1 and PAD2 are of particular interest as both of these genes were identified as being upregulated in the DiD1 cell line that is unable to differentiate efficiently (see

Section 1.8). In order to dissect the mRNA expression of PAD genes throughout the *T. brucei* lifecycle, a Northern hybridisation analysis of several of the PAD genes was performed using RNA from monomorphic slender, stumpy and procyclic form cells.

Digitoxin (DIG) labelled riboprobes had been made previously (Pam Davies, University of Manchester) that were designed to specifically anneal to PAD1, 2, 3 or 5 and 7 mRNA transcripts (PAD 5 and 7 have identical nucleotide sequences in the open reading frame and within the 3' UTR such that it is not possible to distinguish between them) (see Figure 5.1-1). Additionally, a DIG labelled riboprobe made using an amplicon containing the entire PAD2 open reading frame was synthesised to detect transcripts derived from all of the genes within the array simultaneously (see Materials and Methods for details). RNA was resolved on a denaturing agarose gel and stained with ethidium bromide to verify equal loading of the RNA, as determined by visualisation of the ribosomal subunit RNAs (see Materials and Methods for details). The RNA was blotted by capillary action onto positively charged nylon membrane and hybridised with either riboprobe 1, 2, 3 or 5/7 in order to detect the corresponding PAD mRNA in the different lifecycle stages. Hybridised probe was detected by staining with α-DIG followed by chemiluminescence (see Materials and Methods for details).

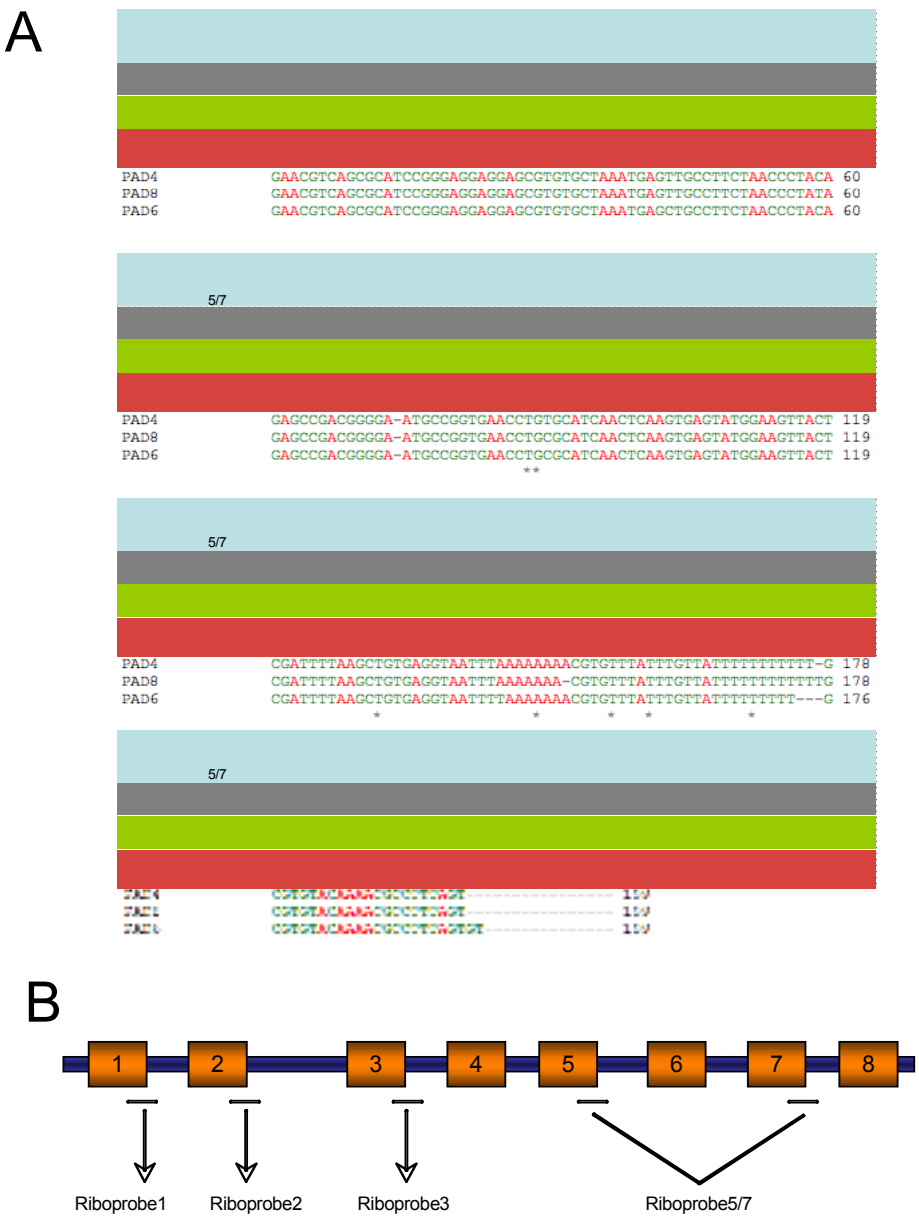


Figure 5.1-1 The riboprobes designed to detect PAD1, PAD2, PAD3, PAD5/7. (A) The sequences of the riboprobes designed to detect specific PAD transcripts were aligned with their target sequences using clustalW (Larkin *et al.* (2007)). In addition, a riboprobe was synthesised using the entire open reading frame of PAD2, that shares enough homology with the other PAD transcripts to detect all of them at once (not shown). (B) The target sequences were derived from the 3' end of the transcripts, including parts of the open reading frame and the predicted 3' UTR.

Figure 5.1-2A shows that the PAD genes have very different expression profiles. PAD1 was highly expressed in stumpy form cells but not in monomorphic slender or procyclic form cells. As discussed in Section 4.1, the Northern blot also shows that PAD1 was highly expressed in the DiD1 cell line (although not as highly as in stumpy form cells), confirming the Macroarray hybridisation data.

In contrast, Figure 5.1-2B shows that PAD2 was highly expressed in procyclic forms, but not in monomorphic slender or stumpy forms. However, as with PAD1, and consistent with the macroarray hybridisation analysis, PAD2 was also highly expressed in the DiD1 cell line.

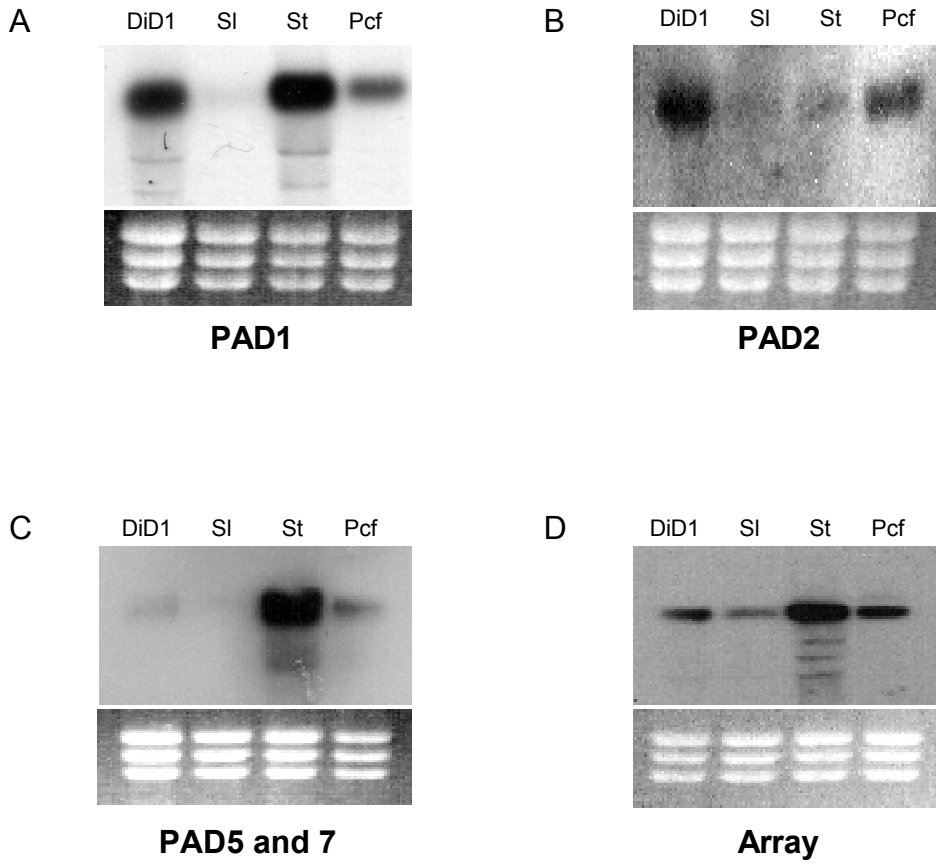
Figure 5.1-2C shows that PAD5 and 7 were highly expressed in stumpy form cells, but not in monomorphic slender or procyclic cells. PAD5 and 7 are thus distinct from PAD1 in that they are not expressed in the DiD1 cell line, despite being upregulated in stumpy form cells and containing similar 3' UTR repetitive elements (see Section 3.3.3)

PAD3 was not detected in either monomorphic slender, or stumpy form lifecycle stages (data not shown) (expression in the procyclic form was not tested), suggesting that either PAD3 is not expressed in any of the lifecycle stages assayed, or that the riboprobe labelling reaction failed. PAD3 was not investigated further. PAD4, 6 and 8 were also not tested for their expression profile, although subsequent microarray analysis indicated that PAD8 was highly upregulated in stumpy form cells (S. Kabani and K. Matthews, unpublished data). Further investigation focused on PAD1 and PAD2 since these genes were enriched in the DiD1 cell line.

Figure 5.1-2D shows the expression of all of the PAD genes simultaneously. It is clear that the highest expression occurs in stumpy form cells, with significant expression in procyclic form cells. PAD array mRNA was detected in monomorphic slender form cells, but not to the extent of the other lifecycle stages. There appeared to be one band predominantly detected; the lower bands probably represent degradation products and were present in the Northern blot corresponding to PAD1 expression. This suggests that the mature PAD mRNA transcripts are in fact of very

similar sizes, conflicting with the predictions from the algorithms that annotate poly-adenylation addition signals (see Section 3.3.1).

Of the 8 genes, PAD1 was thought to have the most interesting expression profile: it was the only member of the array to be implicated in initiation of differentiation through mRNA enrichment in both the DiD1 cell line and in the stumpy form cells. PAD2 was also upregulated in the DiD1 cell line, but as a predominantly procyclic form enriched transcript it was not initially considered to be a primary candidate for a gene having a role in differentiation. PAD5 and 7 were highly expressed in stumpy form cells and were thus both implicated in stumpy form viability or the stumpy to procyclic form transition; however, they were not found to be enriched in the DiD1 cell line and thus they were not analysed in the same depth as PAD1 and PAD2. Moreover, it was expected that much of the functional analysis applied to PAD1 would also be relevant to the other PAD members, thus making the need to investigate every member in depth unnecessary.



E

Ribopobe detection	Slender (monomorphic)	Stumpy	Procyclic	DiD1
PAD1	Barely detectable	++++	+	+++
PAD2	+	+	+++	+++
PAD3	-	-	nd	-
PAD7 and 7	Barely detectable	++++	+	Barely detectable

Figure 5.1-2 A Northern blot analysis of the PAD array mRNA transcripts. A Northern blot was performed to analyse the lifecycle expression of (A) PAD1, (B) PAD2, (C) PAD5 and 7 and (D) the entire array. A non-quantitative summary of the PAD expression data is also shown (E). PAD1, 5 and 7 are upregulated in the stumpy form, whereas PAD2 is upregulated in the procyclic form. The Northern blot that shows expression of the entire PAD array (D) suggests that mRNA levels in the stumpy form are much higher than in either the procyclic form or the slender form.

5.2 Localisation of the PAD1 protein

An important step in elucidating the function of a protein is to determine where the protein localises, and thus where it exerts its biological affect. The bioinformatics evidence concluded that the PAD genes are all transmembrane proteins, but this does not allow determination of anything more specific in terms of sub-cellular location. Possible membranous locations include:

- The cell surface membrane
- The endoplasmic reticulum
- The mitochondrial membrane
- The nuclear membrane
- The Golgi apparatus
- The lysosomal membrane
- The glycosomal membrane

Each of these potential locations has a characteristic immunofluorescence staining profile that can be distinguished using immunofluorescence microscopy and confocal microscopy. Therefore, several strategies were pursued to determine the localisation of the PAD1 protein by immunofluorescence.

5.2.1 Localisation of an ectopically expressed, epitope tagged copy of PAD1

Prior to the availability of an α-PAD1 antibody, localisation of PAD1 was investigated using a cell line expressing an epitope tagged, ectopically expressed copy of PAD1. The location of this ectopically expressed protein can be determined by immunofluorescence utilising the monoclonal BB2 antibody that targets the Ty1 virus-like particle derived epitope tag (Ty tag) (Bastin *et al*. (1996)) incorporated into PAD1 (PAD1-Ty). The advantage of this strategy was that the anti-epitope tag antibody was already available and has been shown to generate good data with little

or no cross-reactivity to other cellular protein (Bastin *et al*. (2001)). However, this strategy suffers from several potential dangers:

- ectopic expression of a protein may overwhelm the intra-cellular transport machinery thus causing its mis-localisation of the resulting protein (for example, an endogenous protein may localise to the endoplasmic reticulum but over-expression of the ectopic copy may cause mis-localisation to the cell surface membrane)
- the high expression of the ectopic copy may cause aberrant folding, causing the protein to be either mis-localised or degraded
- the presence of the Ty epitope tag may interfere with protein folding, targeting signals or with protein-protein interactions, thus causing mis-localisation or degradation

Despite these caveats, this strategy has been used to successfully localise a number of proteins (Hellemond *et al.* (2000)) (Tasker *et al*. (2001)) (Hammarton *et al*. (2003)) and was therefore pursued in the first instance.

5.2.1.1 Generating the DNA constructs pHD451(XhoI) and pHD451(XhoI):PAD1-TY

The DNA plasmid construct pHD451 was designed for inducible expression of gene products in *T. brucei* (Biebinger *et al.* (1999)). Briefly, the gene of interest was cloned between the HindIII and BamHI restriction enzyme digestion sites downstream of an EP procyclin promoter and an EP procyclin 5' UTR. The EP promoter was previously modified to contain an internal tetracycline-inducible repressor (TetR) binding site (i.e. a tetracycline operator sequence) such that, in the presence of the TetR protein, gene transcription is inhibited. Upon addition of tetracycline, TetR forms a complex with tetracycline and no longer binds to the tetracycline operator, thus allowing the initiation of transcription (see Appendix A for a diagram of these DNA constructs). This construct had been previously modified such that cloning a gene into the HindIII and BamHI restriction enzyme sites would incorporate a Ty epitope tag into the 3' end of the open reading frame of the gene, thus the final protein contains a C-terminal Ty epitope tag.

However, cloning PAD1 into this DNA construct without modification was difficult due to the presence of an internal BamHI restriction site within the PAD1 gene. A conventional solution to this problem would be to use a BglII restriction enzyme recognition site within the cloning primers in place of BamHI in order to generate ends compatible with cloning into the vector BamHI site. However, the PAD1 gene also contains an internal BglII site, making this strategy unsuitable. The vector pHD451 was therefore modified to insert a XhoI site between the HindIII and BamHI sites by using reverse complementary oligonucleotide primers (primers #24 and #25) that, when annealed, would form a duplex containing an internal XhoI restriction enzyme site and terminal single stranded nucleotides compatible with cloning into HindIII and BamHI sites (see Figure 5.2-1). This was cloned into pHD451 resulting in the addition of a XhoI restriction enzyme site to the existing HindIII and BamHI site (see Appendix A for a complete list of primer sequences).

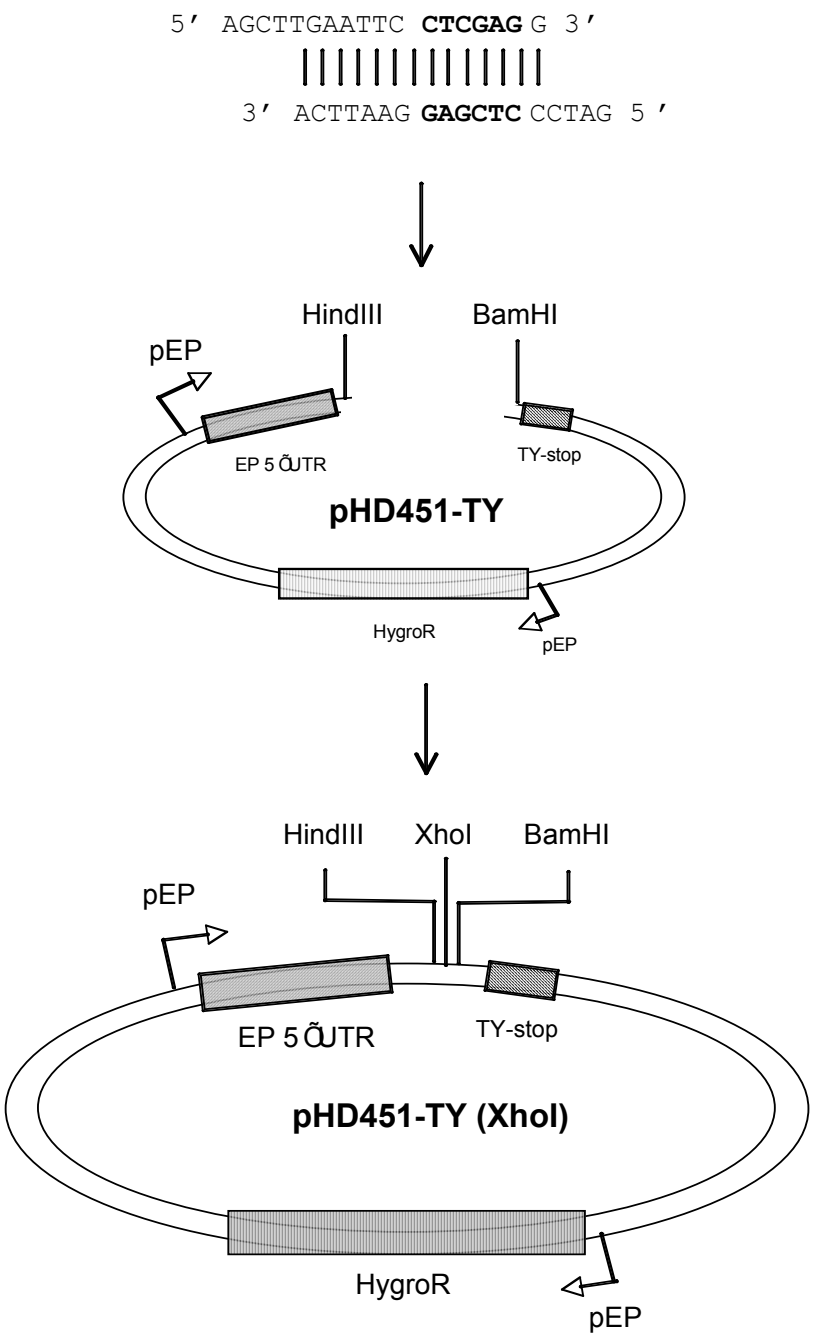


Figure 5.2-1 Construction of pHD451-TY (XhoI). A pair of reverse complementary primers were annealed to form an internal XhoI restriction site and terminal ends complementary to HindIII and BamHI. The annealed primers were ligated to HindIII and BamHI cut pHD451 such that the resulting DNA construct contained HindIII, XhoI and BamHI restriction cloning sites. The XhoI site is shown in bold.

The PAD1 gene was amplified in a polymerase chain reaction from Lister 427 *T. brucei* genomic DNA using primers containing HindIII on the forward primer (primer #26), and XhoI on the reverse primer (primer #27); the stop codon was not included on the reverse primer to allow incorporation of the Ty epitope tag into the C terminus of the protein. The amplicon was digested with HindIII and XhoI and inserted into similarly cut pHD451(XhoI) containing the Ty epitope tag sequence at the 3' end of the multiple cloning site. The resulting recombinant plasmid, pHD451:PAD1-Ty, was then sequenced to verify that there were no errors in the open reading frame of PAD1 incorporated by the DNA polymerase (data not shown).

5.2.1.2 Transfection of trypanosome cells to express PAD1-Ty

The monomorphic bloodstream cell line Lister 427 has previously been transgenically modified (Wirtz *et al.* (1999)) to constitutively express the TetR protein by integration of pHD449 into the ribosomal spacer region of the genome (thus creating the cell line Lister 427-449). This cell line was transfected with NotI linearised pHD451(XhoI):PAD1-Ty DNA construct such that expression of PAD1-Ty was inducible in the presence of tetracycline. Several cell lines were successfully selected after transfection, one of which was investigated for PAD1 localisation.

5.2.1.3 Immunofluorescence microscopy of cells ectopically expressing PAD1-Ty

The cell line Lister 427-449-451:PAD1-Ty was cultured in the presence and in the absence of tetracycline for 2 days prior to being fixed in 3% paraformaldehyde, stained using the BB2 antibody, and counter stained using an α-mouse-FITC conjugated antibody (see Materials and Methods for details) to visualise the cellular location of PAD1-Ty. In addition, a non-transfected cell line was included to control for potential cross-reactivity of the BB2 and α-mouse-FITC antibodies. The resulting slides were examined using immunofluorescence microscopy and images captured using a high resolution digital camera.

The images of the PAD1-Ty induced cells were consistent with cell surface staining (see Figure 5.2-2). There was a faint peripheral concentration of signal around the

cell indicative of cell surface membrane staining, as well as a concentration of signal at the flagellar pocket, consistent with endosomal trafficking and the increased amount of cell surface membrane within this specialised region. In contrast, only background staining was observed when wild type untransfected and non-induced cells were stained, thus confirming that the signal observed when PAD1-Ty is induced is not due to cross-reactivity of the antibody. Interestingly, not all cells in the population exhibited clear staining, suggesting heterogeneity of expression in the population. This has been observed in other studies (Field *et al.* (2004))

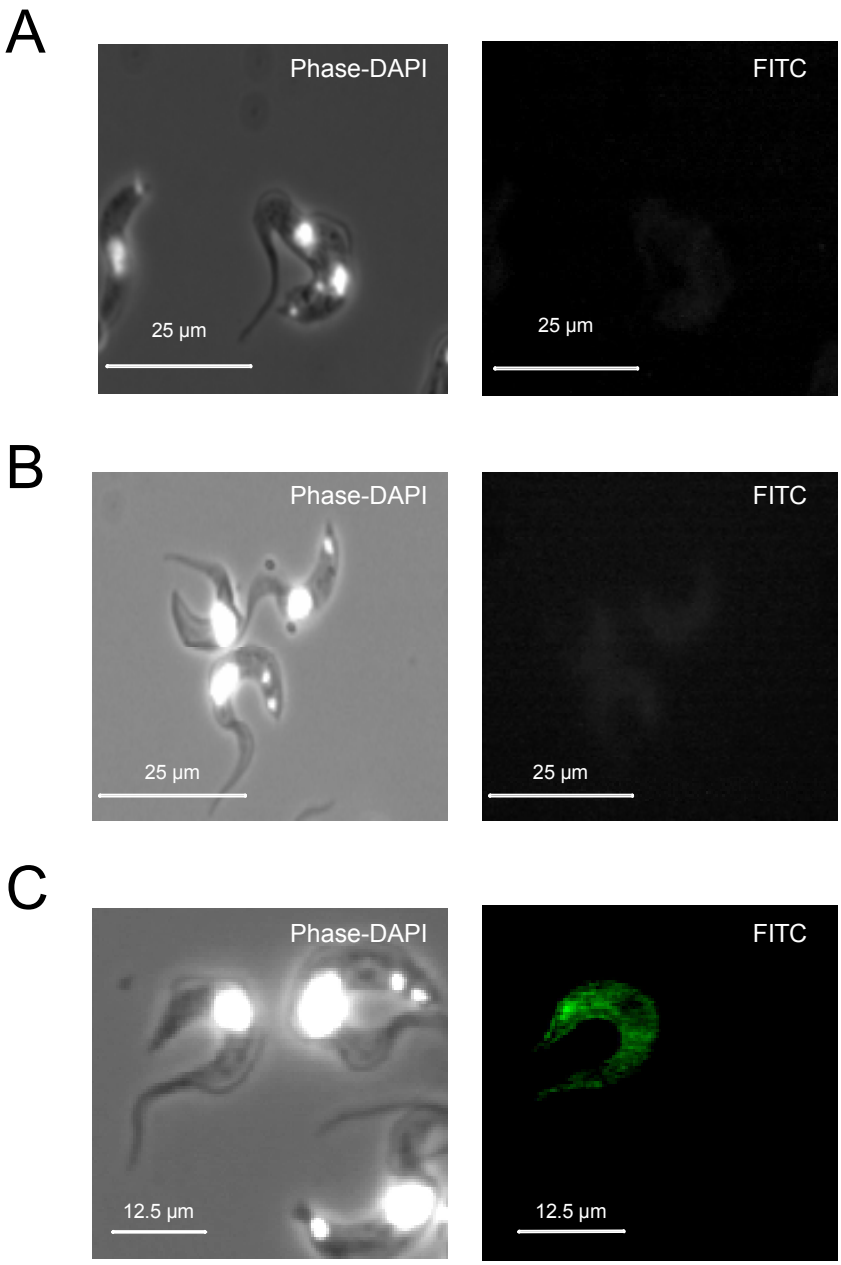


Figure 5.2-2 Localisation of an ectopically expressed epitope tagged copy of PAD1. An ectopically integrated, epitope tagged copy of PAD1 was inducibly expressed under the control of tetracycline and localised by staining with the mouse monoclonal antibody BB2 and an anti-mouse-FITC secondary antibody. (A) Untransfected and (B) non-induced transfected cells showed no staining. (C) Cells induced to express PAD1-Ty showed a staining pattern consistent with localisation to the cell surface membrane, whereby a line around the periphery of the cell, and a concentration of signal around the flagellar pocket (perhaps due to endosomal trafficking), was observed.

5.2.2 Localisation of the endogenous PAD1 protein

Although detection of an epitope tagged gene copy was useful as a first analysis, detection of the endogenous protein was essential to corroborate the localisation of PAD1 at the cell surface membrane There are two strategies to achieve this; (i) integration of an epitope tagged copy of PAD1 into the PAD1 genomic locus such that the regulatory elements of the PAD1 gene remain the same thereby providing expression at physiological levels, and (ii) generating an anti-peptide antibody against the endogenous PAD1. The latter route was chosen for the following reasons:

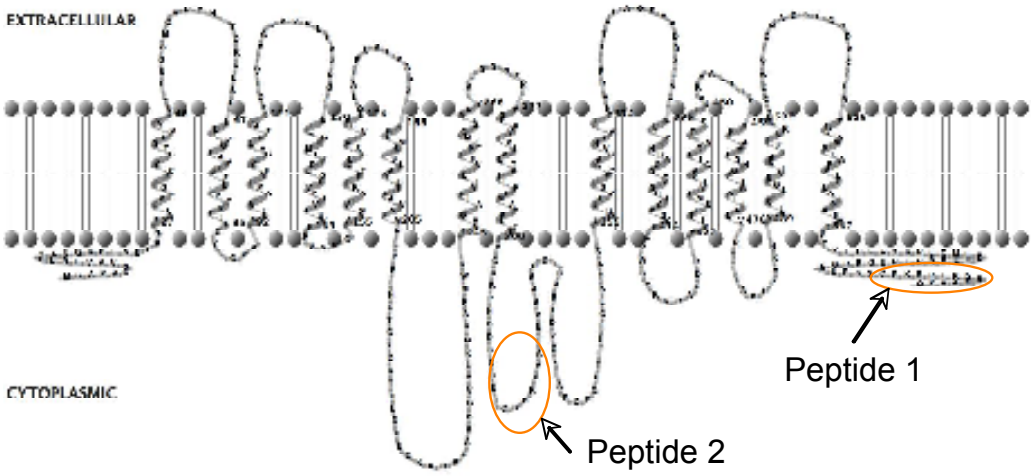
- An α-PAD1 antibody would be a useful tool in dissecting the *T. brucei* lifecycle – for example allowing analysis of wild type slender and stumpy forms
- The Northern analysis suggested that PAD1 was not expressed in the genetically pliable monomorphic cells, whereas generating a transgenic stumpy form that can express an epitope tagged copy of PAD1 is technically difficult
- Visualising the un-modified, endogenous PAD1 protein is a more rigorous strategy to determine localisation
- Presence of an epitope tag may disrupt normal localisation of the expressed PAD1 protein

5.2.2.1 Generating an α-PAD1 antibody

Anti-peptide antibodies can be raised whereby a peptide of 10 – 15 amino homologous to, and specific for, the protein of interest is injected into a suitable animal with an adjuvant over a period of several months. The quality of the resulting polyclonal antibody can vary greatly and can be difficult to predict; for example, background staining, cross reactivity and low target binding affinity are all potential problems that can complicate any subsequent analysis.

Due to the close homology between PAD genes it was necessary to be highly selective when choosing which region of PAD1 to use as a template for a peptide immunogen. In addition, regions that were predicted to form part of trans-membrane regions were excluded because these regions may not be accessible for antibody binding. Two regions were chosen based on these criteria (see Figure 5.2-3). Both were predicted to be intra-cellular such that it was necessary to permeabilised the cell prior to staining (see Figure 5.2-3A), and were from regions of the PAD protein distinct from other members of the family (see Figure 5.2-3C). In addition, a BLASTP search using the peptide sequences as a query against the completed trypanosome genome database gave no 'hits' such that risk of antibody cross-reactivity to other trypanosome proteins was minimised (data not shown). The antibodies were ordered from Eurogentec using their "DoubleXP" programme: the synthetic peptides were inoculated into rabbits with an adjuvant 4 times over a period of 4 months and subsequently affinity purified against the corresponding peptide to prevent non-specific antibody contamination.

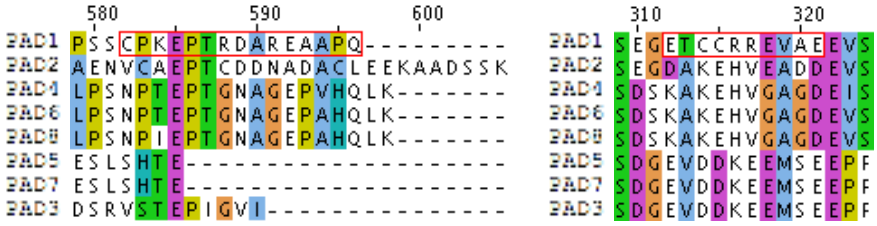
A



B

Antibody	*Amino acid sequence*
PAD1 SPECIFIC1 (15 residues)	H_2N - CPK EPT RDA REA APQ - COOH
PAD1 SPECIFIC2 (10 residues)	H_2N - ETC CRR EVA E - COOH

C



PAD1 specific #1 PAD1 specific #2

Figure 5.2-3 Designing an anti-peptide antibody specific to PAD1. (A) The positions of the sequences used to design the immunogenic peptides are indicated by orange circles on the PAD1 secondary structure. (B) The sequences of the peptides are shown. (C) A clustalW alignment of the PAD proteins showing the two regions that the peptides were designed against (the peptide sequence is indicated by the red square). The left alignment represents αPAD#1, and the right sequence represents αPAD #2.

5.2.2.2 Localisation of the endogenous PAD1 using an α-PAD1 antibody

Early attempts to validate the α-PAD1 antibodies by Western blot were unsuccessful due to technical issues associated with the detection of membrane associated proteins (see Section 5.3 for more details). For this reason, the antibodies were initially validated by immunofluorescence microscopy, and were subsequently used successfully in Western blot with appropriate negative controls to confirm that they did not cross react other, slender cell, proteins (see Section 5.4.1).

An analysis of the ectopic expression of PAD1-Ty indicated that PAD1 was probably localised to the cell surface membrane (see 5.2.1). To investigate the endogenous protein, stumpy form trypanosomes were purified from a mouse infection and reacted with α-PAD1 antibodies in an immunofluorescence analysis (see Materials and Methods for details). Images were analysed by confocal microscopy to obtain the maximum resolution of cellular structures (see Materials and Methods for details).

Figure 5.2-4 shows that the staining was consistent with localisation to the cell surface membrane. There was a generalised staining across the cell, but a clear line of intense labelling around the periphery of the cell. There did not appear to be staining characteristic of the nucleus, endoplasmic reticulum, Golgi apparatus or any other of the intracellular organelles. Note that, slender cells did not stain with the α-PAD1 antibody (for further analysis of the stumpy cell specific staining, see Section 5.4.2, Figure 5.4-1) providing a rigorous internal control for non-specific antibody binding of both the primary and secondary antibodies.

Thus, both the from the ectopic, epitope tagged copy of PAD1 and the endogenous copy of PAD1 demonstrated cell surface membrane localisation.

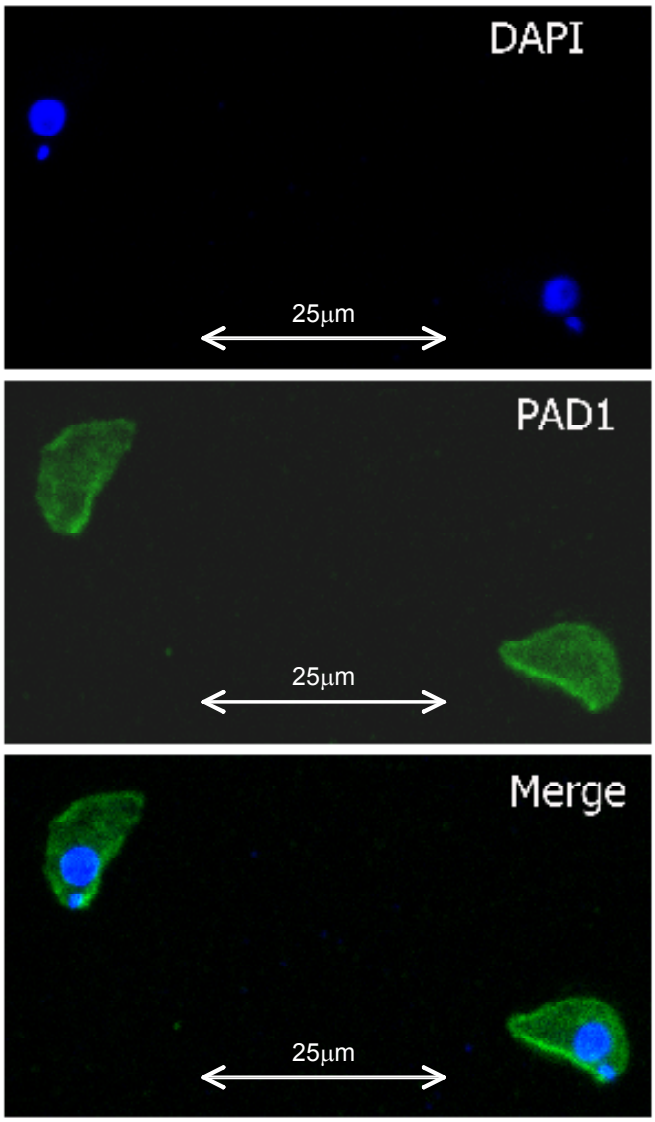


Figure 5.2-4 The PAD1 protein is localised to the cell surface membrane. Stumpy form trypanosomes were fixed with 3% paraformaldehyde and permeabilised with 0.1% TritonX prior to settling on poly-lysine coated glass slides. The cells were stained with the α-PAD1 antibody #1 and visualised using confocal microscopy. PAD1 appears to localise to the cell surface membrane of the trypanosome. DAPI staining is also shown to visualise the kinetoplast and nucleus.

5.3 Optimising detection of PAD1 by Western analysis using an ectopically expressed, epitope tagged copy of PAD1

Transmembrane proteins can be technically challenging to detect on Western blots due to their hydrophobic properties and their tendency to precipitate. Initial attempts to detect the Ty epitope tagged copy of PAD1 by Western blot were unsuccessful, despite the fact that this protein was clearly detectable by immunofluorescence (see Section 5.2.1.3). Therefore, it was decided to optimise the Western detection of PAD1 using the Ty-specific BB2 antibody on cells expressing PAD1-TY instead of the α-PAD1 antibody because:

- The BB2 antibody has excellent binding affinity and has been used successfully in Western analysis previously with little or no cross reactivity (Bastin *et al*. (1996))
- Availability of the BB2 hybridoma line meant that an unlimited supply of the antibody was available
- The specificity of the newly made α-PAD1 antibodies was not well established

A number of variables were considered before any band became detectable by Western analysis: the pH and type of transfer buffer; the voltage used in semi-dry blotting; the time of semi-dry transfer and the amount of protein loaded (data not shown). However, when the temperature of denaturation prior to loading the protein on to the SDS polyacrylamide gel was varied, there was an inverse correlation between the temperature of denaturation and the detection of PAD1-Ty (see Figure 5.3-1). When the protein was not heated at all, a clear band was detected at approximately 55 kDa. This was lower than the 67 kDa that was predicted for PAD1, which is not unusual for hydrophobic proteins. However, it is interesting to note that when the denaturation temperature was raised, the PAD1-Ty protein appeared to resolve at a higher molecular mass (albeit much less intensely), suggesting that the aberrant migration at low temperature might be due to its

secondary structure. Thus, it appears that at higher temperatures PAD1 precipitates precluding detection either due to inability of the protein to enter the gel, unavailability of antigenic sequence to the antibody, or both.

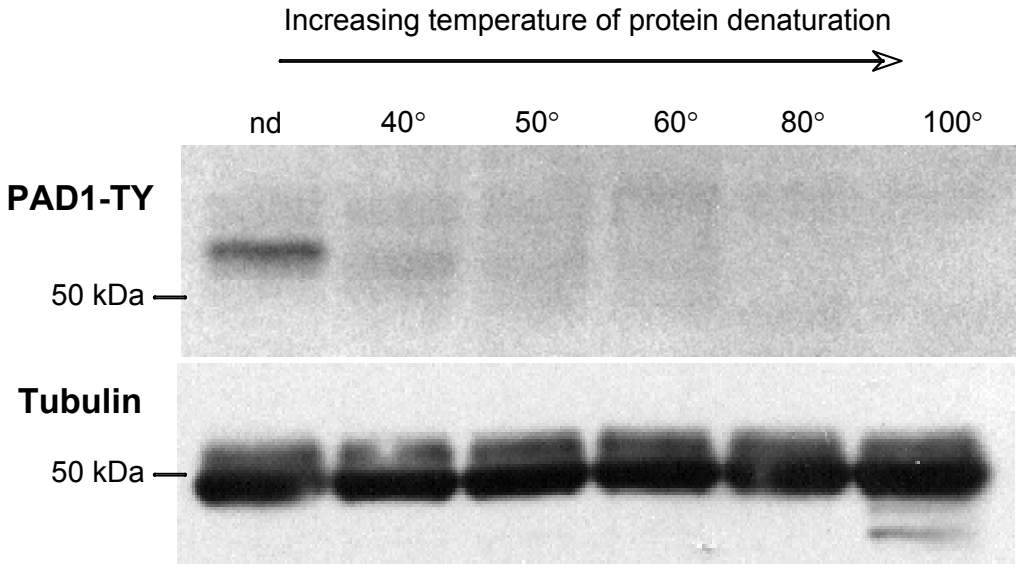


Figure 5.3-1 The effect of denaturation temperature upon PAD1-Ty detection by Western blot. Proteins from cells ectopically expressing PAD1 with a Ty epitope tag were either not heat denatured (nd) or denatured at 40°, 50°, 60°, 80° or 100° after cell lysis in Laemmli sample buffer. The protein was analysed by SDS-PAGE and Western analysis. PAD1-Ty was only detected clearly when the protein was not heat denatured, suggesting that the protein precipitates at higher temperatures precluding detection. High molecular weight forms of the protein were detected upon heating, although they were far less intense. The upper blot is an exposure of PAD1-Ty using the BB2 antibody, and the lower blot is the same membrane reacted with an antibody recognising trypanosome α-tubulin (Woods *et al.* (1989)).

The Western blot conditions optimised for PAD1-Ty detection using the BB2 antibody were transposed to analysis of endogenous PAD1 detected using the α-PAD1 antibody. Further optimisation whereby the proteins were resolved at lower voltages, using chilled buffers to maintain a low temperature, and transferred onto PVDF membrane by a wet blotting system, further enhanced sensitivity, such that detection was improved from an exposure time of several hours to less than 5 seconds (see Materials and Methods for the fully optimised Western detection protocol).

5.4 PAD1 is a stumpy form specific protein

Investigations into blood form trypanosome differentiation have been hindered by the absence of an absolute marker discriminating slender from stumpy form cells. Historically, morphology was the preferred way of distinguishing the lifecycle stages of bloodstream trypanosomes (Bruce (1911)) (Robertson *et al.* (1913)), but this was complicated by the presence of intermediate forms that share characteristics of both stumpy and slender forms within a natural pleomorphic population. Subsequently, it was shown that stumpy cells and not slender cells have a partially active mitochondrion and show mitochondrial diaphorase activity, as determined by the formation of formazon deposits when cells were stained with $NADH_2$ and nitro-blue tetra-zolium salt (Vickerman (1965)). This was extended (Tyler *et al.* (1997)) using an antibody specific to dihydrolipoamide dehydrogenase (DHLADH) and other active mitochondrial enzymes in a Western and immunofluorescence analysis of mitochondrial marker expression during differentiation between slender and stumpy forms. However, using mitochondrial proteins to distinguish stumpy form cells is not ideal as eukaryotic cells are known to upregulate many mitochondrial proteins in response to oxidative stress or apoptotic conditions (Giffin *et al.* (1986)) (Giffin and McCann (1989)) (Lin and Beal (2006)) (Sapienza *et al.* (2008)). It is therefore difficult to distinguish between conditions that induce cells to differentiate to the stumpy form and those that simply induce a stress response. The development of a stumpy form surface marker unrelated to the stress response would therefore be an invaluable tool in the molecular analysis of the *T. brucei* lifecycle. The next series of

experiments set out to investigate the potential for PAD1 to act as a marker for stumpy form cells.

5.4.1 Western analysis

In order to determine the protein expression profile of PAD1, protein from monomorphic slender, stumpy and procyclic form cells was resolved on an SDS polyacrylamide gel for Western analysis using the previously optimised conditions. The Western analysis of PAD1 detected a strong band resolving at approximately 55 kDa (see Figure 5.4-1A), approximately 11 kDa smaller than is predicted from the primary amino acid sequence. Nonetheless, there seems little doubt that this band is in fact PAD1 and the disparity is probably due to its hydrophobicity or secondary structure; cleavage of the protein is unlikely as this would probably disrupt the structure of the mature protein and prevent proper localisation to the cell surface membrane. No band was detected in either the monomorphic slender or procyclic form lanes (Figure 5.4-1A). Consistent with the mRNA expression data (see Section 5.1), the Western data therefore demonstrates that PAD1 is a stumpy form specific protein.

5.4.2 Immunofluorescence

If PAD1 is a stumpy-specific protein, then it should be possible to distinguish between individual slender and stumpy cells within a population by immunofluorescence staining with α-PAD1. It would be expected that cells with a stumpy morphology would stain positive for PAD1, whereas cells with a slender morphology would be negative for PAD1. Furthermore, where morphology is ambiguous, it would be expected that a PAD1 positive cell would be cell cycle arrested in G1 (termed G0) and would therefore have 1 nucleus and 1 kinetoplast (1K1N), and that cells that are undergoing division (2 kinetoplasts and 1 nucleus, or 2 kinetoplasts and 2 nuclei) would be negative for PAD1. These karyotypes can be simply assayed by staining cells with 4',6-diamidino-2-phenylindole (DAPI) followed by visualisation by fluorescence microscopy.

To analyse the expression of PAD1 at the single cell level, stumpy form cells were purified from a mouse infection (see Materials and Methods) and fixed with methanol prior to staining sequentially with α-PAD1 and α-rabbit-cy3 antibodies. Methanol is unsuitable for localisation studies because the intra-cellular compartments become disrupted, but was used as a fixative in this analysis for the preservation of cellular morphology to distinguish between slender and stumpy cells. Cells from a late stage parasitaemia were used so that the morphological analysis would be less complicated by the presence of intermediate forms within the population.

Consistent with PAD1 being a stumpy form specific protein, essentially all cells with a stumpy morphology stained positive for PAD1 (<99%), and the few cells that were found with a slender morphology were overwhelmingly negative for PAD1 (93%, N = 154 from a total of over 5000 cells) (Figure 5.4-1B).

Interestingly, when the PAD1 positive cells with a slender morphology are analysed, approximately half are 1K1N, half are 2K2N, and none are 2K1N; one might therefore speculate that the commitment to the stumpy form occurs in the 2K2N cell cycle stage, after which cells go into G1 arrest and then become morphologically stumpy. However, only 11 cells (1.8%) from a total of >600 were found that were both PAD1 positive and morphologically slender, thus it is not possible to draw firm conclusions from this analysis.

Two representative images that show PAD1 negative, dividing slender cells with adjacent G1 arrested PAD1 positive stumpy form cells are also shown (see Figure 5.4-1C).

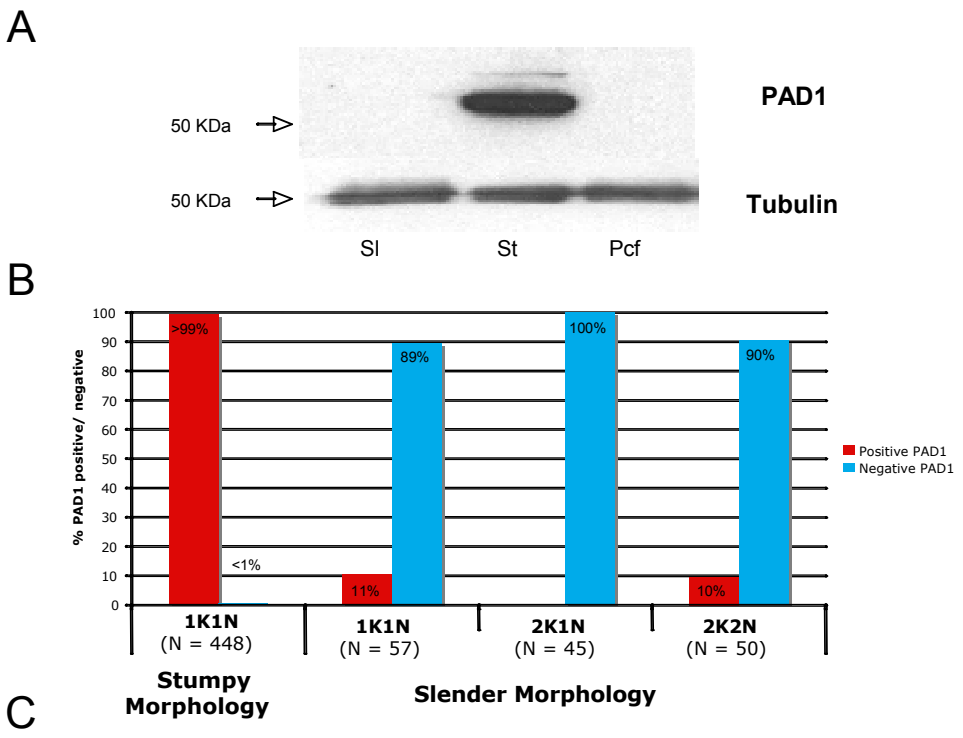


Figure 5.4-1 PAD1 is a stumpy form specific protein. (A) Slender, stumpy and procyclic form cell lysate was resolved by SDS-PAGE and probed with the anti-PAD1 #1 antibody in a Western analysis. PAD1 is undetectable in either slender or procyclic form cells, but strongly expressed in the stumpy form. (B) Pleomorphic trypanosomes were stained with the α-PAD1 antibody and DAPI. Two PAD1 positive stumpy cells and two PAD1 negative dividing slender cells are shown as representative images. (C) Cell counts were performed to determine the proportion of morphologically slender and stumpy cells that are positive for PAD1, and this was related to cell cycle position by counting the number of nuclei and kinetoplast. There is a strong correlation between stumpy morphology and PAD1 staining, and a negative correlation between slender morphology and PAD1 staining.

5.4.3 Flow cytometry analysis

Flow cytometry allows the quantitative protein detection of a large number of individual cells to be determined. Cells are fixed with a suitable fixative (such as formaldehyde/ glutaraldehyde, paraformaldehyde or methanol), stained with a primary antibody and counterstained with a fluorochrome-conjugated secondary antibody. Cells containing target protein-antibody complex are detected by measuring the emission from excitation of the fluorochrome with a laser.

To assess the ability of PAD1 to be used to distinguish between slender and stumpy form cells by flow cytometry, monomorphic slender and stumpy form cells were fixed and stained with α-PAD1 #1 and α-rabbit-FITC for flow cytometry analysis (see Materials and Methods for details). The resulting data were analysed for fluorescence in the FL1 emission channel using FlowJo (Tree Star, Inc., Ashland, OR).

Stumpy form cells form a distinct peak, with higher fluorescence on the histogram than monomorphic slender form cells (see Figure 5.4-2). This allows us to conclude that stumpy form cells express PAD1 at higher levels than monomorphic slender form cells, supporting the hypothesis that PAD1 is a stumpy specific protein.

Furthermore, the development of a protocol that allows stumpy form cells to be distinguished form slender cells and analysed by flow cytometry and fluorescence activated cell sorting (FACS) opens the possibility for a number of other experiments. For example, the population kinetics of the slender to stumpy transition within a parasitaemia can be modelled using either using the endogenous PAD1 protein, or by fusing the PAD1 3' UTR to a reporter gene such as green fluorescent protein (GFP). Additionally, with a high throughput assay for stumpy forms it should be possible to screen for the, as yet unidentified, stumpy induction factor (SIF) (see Chapter 7 for further discussion).

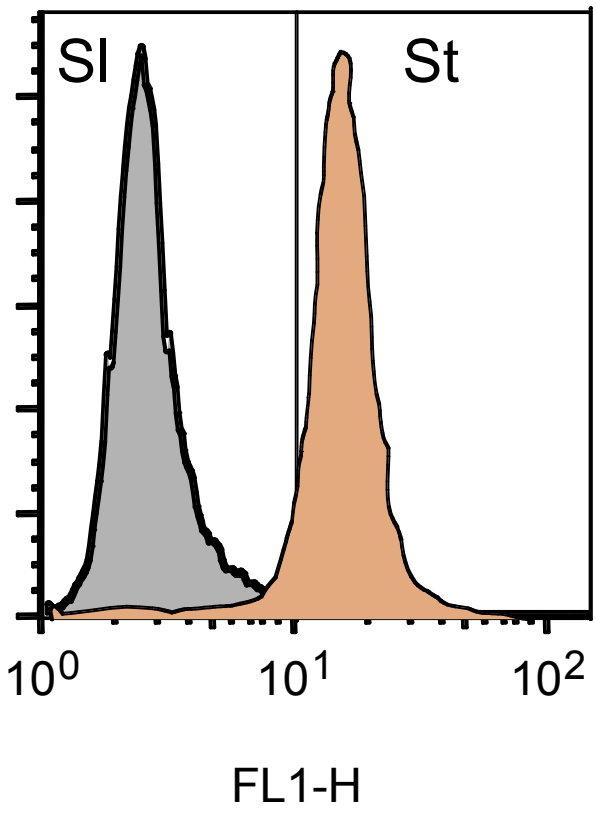


Figure 5.4-2 Slender and stumpy form populations can be distinguished by flow cytometry. Monomorphic slender cells and stumpy form cells harvested from mice were fixed in 1% paraformaldehyde and stained with the α-PAD1 antibody in permeabilisation solution, and counterstained with α-rabbit-FITC. The stumpy population (St) clearly fluoresces in the FITC channel with greater intensity than the monomorphic slender population (Sl). Thus slender and stumpy form populations can be distinguished by flow cytometry and FACS analysis.

In conclusion, the Western, immunofluorescence and flow cytometry data demonstrate that PAD1 is the first unambiguous stumpy form molecular marker. As a non-mitochondrial protein it is not subject to oxidative-stress induced upregulation and other non-specific affects that DHLADH and other mitochondrial enzymes may be affected by. Furthermore, its localisation to the cell surface membrane means that it is possible to clearly determine whether a cell is PAD1 positive. This provides an invaluable tool for analysing the differentiation between the slender, stumpy and procyclic form lifecycle stages. Additionally, the use of PAD1 as a stumpy form specific protein allows the verification of chemicals that have been reported to induce trypanosomes to differentiate to the stumpy form. It would be of interest to further dissect the timing of expression of different stumpy markers to determine at what point a cell commits to becoming a stumpy form (see Chapter 7 for further discussion).

5.5 Investigating PAD2 expression

The elevated expression of PAD2 mRNA in DiD1 cells meant that it was of interest to investigate this protein further.

5.5.1 Designing antigenic peptides to generate α-PAD2 antibodies

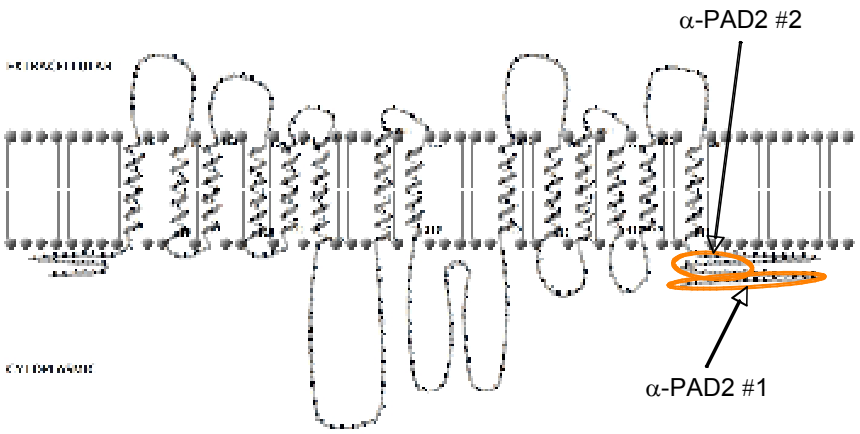
In order to determine the protein expression profile of PAD2, two antigenic peptides were designed that would generate polyclonal antibodies when rabbits were immunised. As with the α-PAD1 antibodies, regions of the PAD2 protein were chosen such that the peptide was not predicted to be within the N-terminal signal anchor, or any of the 14 transmembrane helices. Additionally, a protein sequence alignment was generated in order to identify regions that were unique to PAD2 and therefore not represented in any of the other PAD proteins. Due to these highly stringent requirements, only 1 region at the carboxyl terminus was found that was suitable for antigenic peptide design.

Two peptides were chosen based on the amino acid sequence from the carboxyl terminus of PAD2. Peptide #1 consists of 15 residues at the extreme carboxyl terminus of PAD2, and peptide #2 consists of 10 residues near the carboxyl terminus

(see Figure 5.5-1A and B). An alignment shows that these peptide do not show significant homology to any of the other PAD proteins (see Figure 5.5-1C). Additionally, a BLAST search using both of these peptides as query sequences did not give any matches against the *Trypanosoma brucei* database (data not shown).

The peptide sequences were submitted to Eurogentec for immunisation into rabbits and antibody generation using their 28 day 'Speedy Double XP Programme'. The antibodies were affinity purified using the immunogenic peptides by Eurogentec.

A



B

Antibody	*Amino acid sequence*
α-PAD2 #1 (10 residues)	H_2N Š CAN DAC LEE KAA DSS K Š COOH
α-PAD2 #2 (15 residues)	H_2H Š IKE AED NQT NAE NVC Š COOH

C

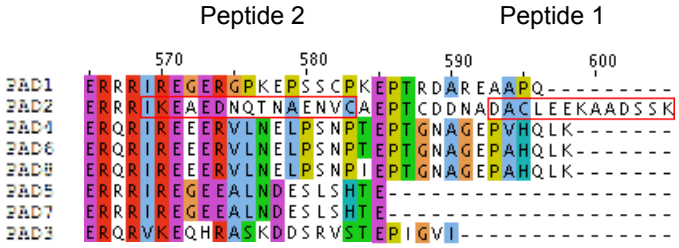


Figure 5.5-1 Designing the antigenic peptides for the α-PAD2 antibodies. (A) The positions of the sequences used to design the peptides are indicated by orange circles on the PAD2 secondary structure. (B) The sequences of the peptides are shown. (C) A ClustalW alignment of the PAD proteins showing the region that the peptides were designed against (the peptide sequences are indicated by the red square). The sequences were chosen such that the resulting antibodies should only bind to PAD2.

5.5.2 A Western blot analysis to determine PAD2 expression throughout the trypanosome lifecycle

Cell lysate protein from monomorphic slender, stumpy and procyclic forms was prepared and analysed by Western blot with the α-PAD2 antibodies using conditions optimised for PAD1 detection (see Materials and Methods for details).

Figure 5.5-2 shows that a 57 kDa band was detected in both the stumpy form and the procyclic form lanes, with the procyclic form band showing approximately 3 fold elevated expression in comparison with the stumpy form. Several strong bands at approximately 60 kDa were detected in Lister 427 monomorphic slender cells by the α-PAD2 #1 antibody, and the same bands but much weaker (although equally diffuse) were detected by α-PAD2 #2. These bands were almost certainly due to the α-PAD2 antibodies cross reacting with VSG expressed by these monomorphic cells. Indeed, despite their obviously high expression, these bands were not detected by an α-array antibody that was reactive against all of the PAD members (see Section 5.6.2).

Thus, it would appear that PAD2 is expressed in stumpy forms, and approximately 3 to 4 fold higher in procyclic forms. Despite the apparent Lister 427 VSG cross-reactivity, no convincing expression was detected in the monomorphic slender cell lysate.

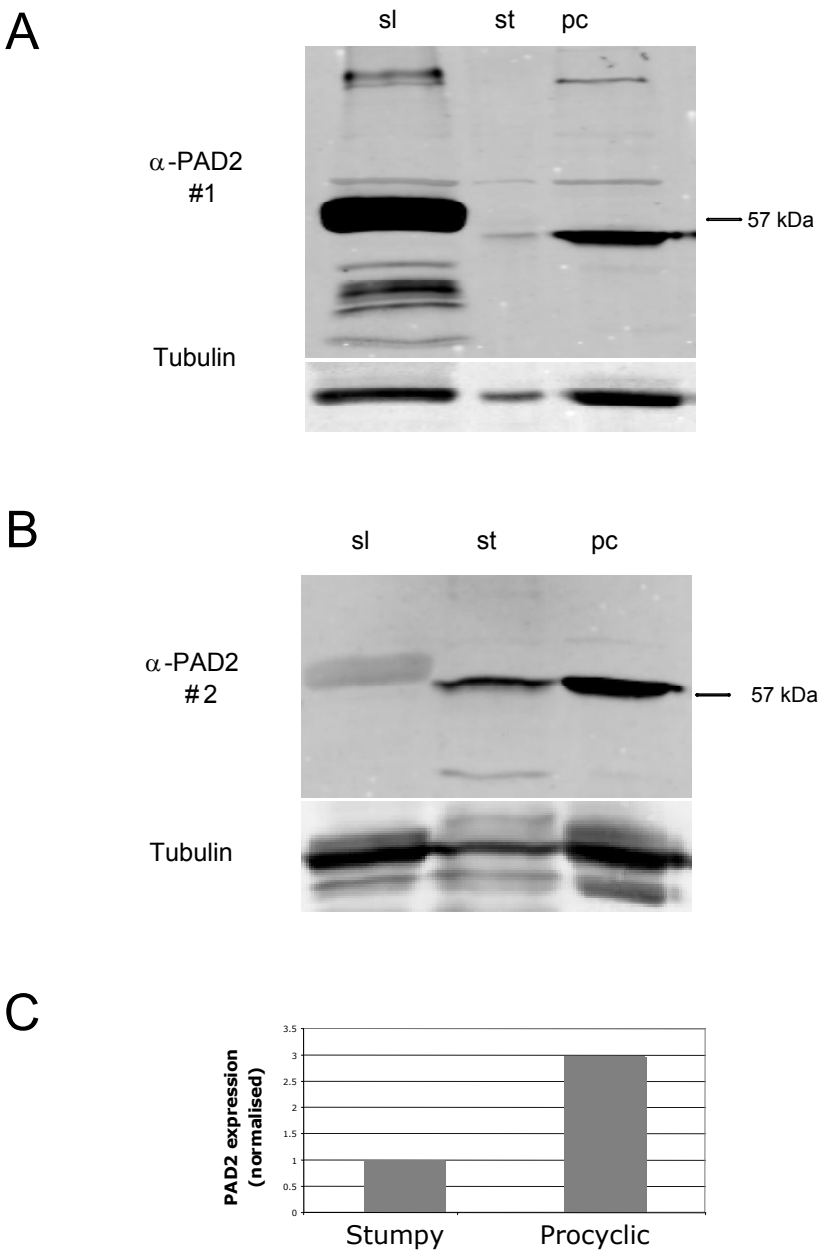


Figure 5.5-2 The lifecycle protein expression of PAD2. Protein from monomorphic slender, stumpy and procyclic form trypanosomes was analysed by Western blot for PAD2 expression using (A) α-PAD2 #1 and (B) α-PAD2 # 2. (C) The signal from panel B was quantified using the LI-COR odyssey system and expressed relative to the tubulin loading control. Procyclic forms express PAD2 at levels approximately 3 fold that of stumpy forms. Monomorphic slender forms do not appear to express PAD2. Bands detected in the 'sl' lane are probably cross-reactivity with VSG (which migrates at 60 kDa).

5.6 Analysing the lifecycle expression of the entire PAD family

As described (see Section 5.1), PAD1 is a member of an 8 gene cluster with differential RNA expression through the *T. brucei* lifecycle. It was necessary to concentrate on only some of the PAD members to perform an in depth analysis, and PAD1 and PAD2 were chosen for the reasons discussed. However, PAD5 and 7 mRNA transcripts were also shown to be upregulated in stumpy form cells, and the PAD2 mRNA transcript was shown to be upregulated in the DiD1 cell line as well as in the procyclic form. Thus it appeared that this was an interesting array of genes and it was decided to investigate other members to some extent, in addition to PAD1 and PAD2.

5.6.1 Generating antigenic peptide antibodies recognising all PAD array members

The highly homologous nature of the PAD array, while making some types of analysis more difficult (for example Northern analysis and RNAi mediated transcript ablation), had the advantage that it is comparatively simple to design an antibody that detects multiple members of this cluster. To achieve this, the PAD protein sequences were aligned by clustalW (see Figure 5.6-1C, and Appendix A for a full Multalin alignment) and one peptide sequence was selected that was common to the entire array, and one was chosen that was predicted bind to some members with higher affinity, possibly allowing specific identification of any resulting bands (see Figure 5.6-1). The requirement for these regions were:

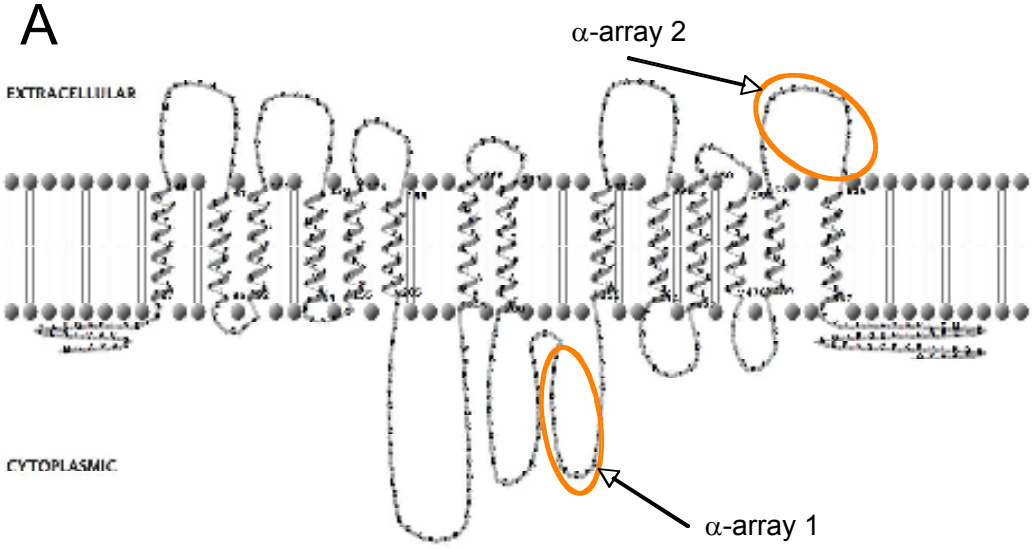
- At least 15 – 20 amino acids where there was little or no heterogeneity in the PAD alignment
- They were predicted to be either extra cellular or intracellular (i.e. not membranous) to facilitate antibody binding
- They should not form part of the N-terminal signal anchor – although signal anchors are not predicted to be cleaved, it is possible that the PAD proteins

have a cryptic cleavage site resulting in the antibody target sequence not forming part of the mature protein

- A BLAST search identifies no other matches in the *T. brucei* database, thus minimising the risk of cross-reactivity (data not shown)

The peptide sequences were sent to Eurogentec for synthesis and inoculation into rabbits followed by affinity purification against the antigenic peptide using their "DoubleXP" programme.

A



B

Antibody	*Amino acid sequence*
α-array #1 (15 residues)	H_2N - VET DVD YIA PQF QET - COOH
α-array #2 (15 residues)	H_2N - TQQ ADK LGQ DVC TER - COOH

C

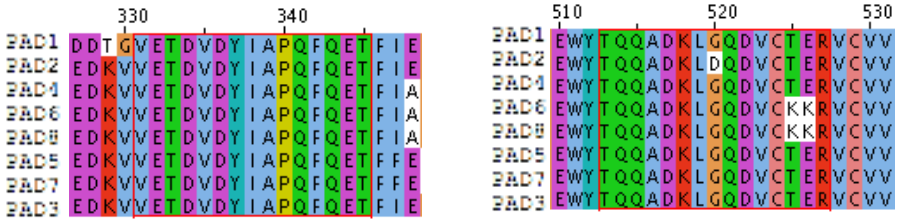


Figure 5.6-1 Designing the antigenic peptides for the α-array antibodies. (A) The positions of the sequences used to design the peptides are indicated by orange circles on the PAD1 secondary structure. (B) The sequences of the peptides are shown. (C) A ClustalW alignment of the PAD proteins showing the two regions that the peptides were designed against (the peptide sequence is indicated by the red square). Anti-array #1 (left) is predicted to bind to all of the PAD proteins equally well, whereas anti-array #2 (right) may bind to PAD6 and 8 with lower affinity.

5.6.2 Expression of unidentified PAD proteins through the trypanosome lifecycle

Given that PAD1 mRNA and protein is stumpy form specific, and that PAD2 mRNA transcript is enriched in the procyclic form, it was of interest to determine the relative levels of PAD array protein expression in different lifecycle stages, and also to determine if any of the PAD proteins were expressed in monomorphic slender cells. This was possible because the Northern blot analysis using a riboprobe to detect all of the PAD gene mRNA transcripts showed that there was at least some PAD gene RNA expression in the monomorphic slender form (see Section 5.1, Figure 5.1-2).

Cell lysate was made from monomorphic slender, stumpy and procyclic form cells and analysed by Western blot with the α-array #1 and α-array #2 using the protocol optimised for PAD1 detection (see Materials and Methods). The introduction of the LI-COR infrared system (Weldon *et al*. (2008)) into the laboratory allowed a relative quantitation of the resulting signal in different lifecycle stages that had previously been unavailable. Thus, infrared fluorochrome conjugated secondary antibodies were used in place of horse-radish peroxidase (HRP) conjugated antibodies. Additionally, the use of fluorochromes with non-overlapping emission spectra allowed co-staining with an α-tubulin loading control to normalize the signal from the α-array antibodies for relative sample loading.

Figure 5.6-3 shows that when the α-array #1 antibody was used as a primary antibody in a Western blot analysis of the different lifecycle cell lysates, no PAD protein was detected in monomorphic slender cells, a 55 kDa and a 57 kDa band was detected in the stumpy cell lysate, and a 57 kDa band was detected in the procyclic form cell lysate. Similar results were obtained using the α-array #2 antibody. A faint band resolving at approximately 54 kDa was also detected by α-array #2 in procyclic cell lysate, but this probably represents a degradation product of the 57 kDa band, and was not detected in all procyclic form cell lysates (data not shown). Perhaps interestingly, the 57 kDa band detected by α-array #1 is much fainter than the 55 kDa band, whereas the 2 bands are of approximate equal intensity when the α-

array #2 antibody is used. This may represent differential affinity to different PAD proteins (for example, the α-array #2 antibody may bind to different member with differing affinities; Figure 5.6-1C) between the 2 antibodies, but probably reflects protein sample variation.

Thus, it would appear that 2 different PAD proteins are expressed in the stumpy form, and 1 PAD protein is expressed in the procyclic form. It is possible that several PAD proteins contribute to either the 55 kDa or the 57 kDa band, but it would seem unlikely that they would resolve so closely as to be indistinguishable on a 10% poly-acrylamide gel. No PAD proteins were detected in the monomorphic slender cells, confirming that the bands detected by the α-PAD2 antibody were indeed due to VSG cross reactivity; note that this was in apparent contrast with the Northern blot data (see Section 5.1), that shows that there is PAD mRNA expression in these cells. This may reflect a mechanism of translational control specific to slender cells.

Further experiments were performed to identify which of the PAD proteins corresponded to each band.

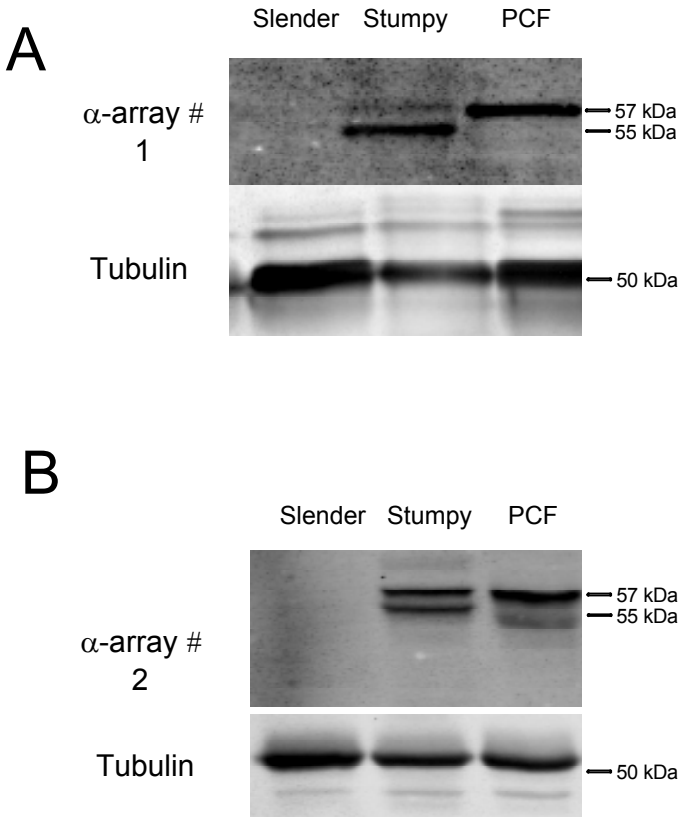


Figure 5.6-2 The expression of PAD proteins though the *T. brucei* lifecycle. Cell lysate from different lifecycle stages of *T. brucei* were analysed by Western blot (A) α-PAD array #1 and (B) #2 antibodies. Both antibodies detected a 55 kDa and 57 kDa band in stumpy form cell lysate and a 57 kDa band in procyclic form ell lysate. A 54 kDa band is also detected in procyclic form cell lysate by α-array #2, but this probably represents a degradation product of the 57 kDa band.

5.6.3 Identification of the bands detected by the α-array antibodies

To identify which of the PAD proteins the bands detected by the α-array antibodies corresponded to, cell lysate from stumpy and procyclic form was resolved by SDS-PAGE and transferred onto PVDF membrane by wet blotting. Each of the lanes was cut in half longitudinally and 1 half was analysed by Western blot using either α-PAD1 or α-PAD2, and the other half of the lane was analysed using α-array #2. α-array #2 was chosen as this gave the best signal for both bands in the stumpy lane (Figure 5.6-2). Post-staining, the lanes were precisely aligned and analysed using the LI-COR Odyssey infrared imaging system (see Materials and Methods for details).

Figure 5.6-3 shows that the 55 kDa band in the stumpy cell lysate was confirmed as PAD1, and the 57 kDa band in both the stumpy form and the procyclic form was identified as PAD2.

These data do not exclude the possibility that other PAD proteins resolved at the same size and also contribute to the signal detected by the α-array antibodies. It is worth noting that, although PAD5/7 RNA expression profiles suggest stumpy specific expression, a third band was not detected in the stumpy form protein; PAD5/7 were predicted to be smaller than PAD1 and PAD2 (0.8 kDa and 1.7 kDa respectively), and may therefore be expected to resolve differently. Therefore, it may be that PAD5 and PAD7 are not expressed at the protein level under these conditions.

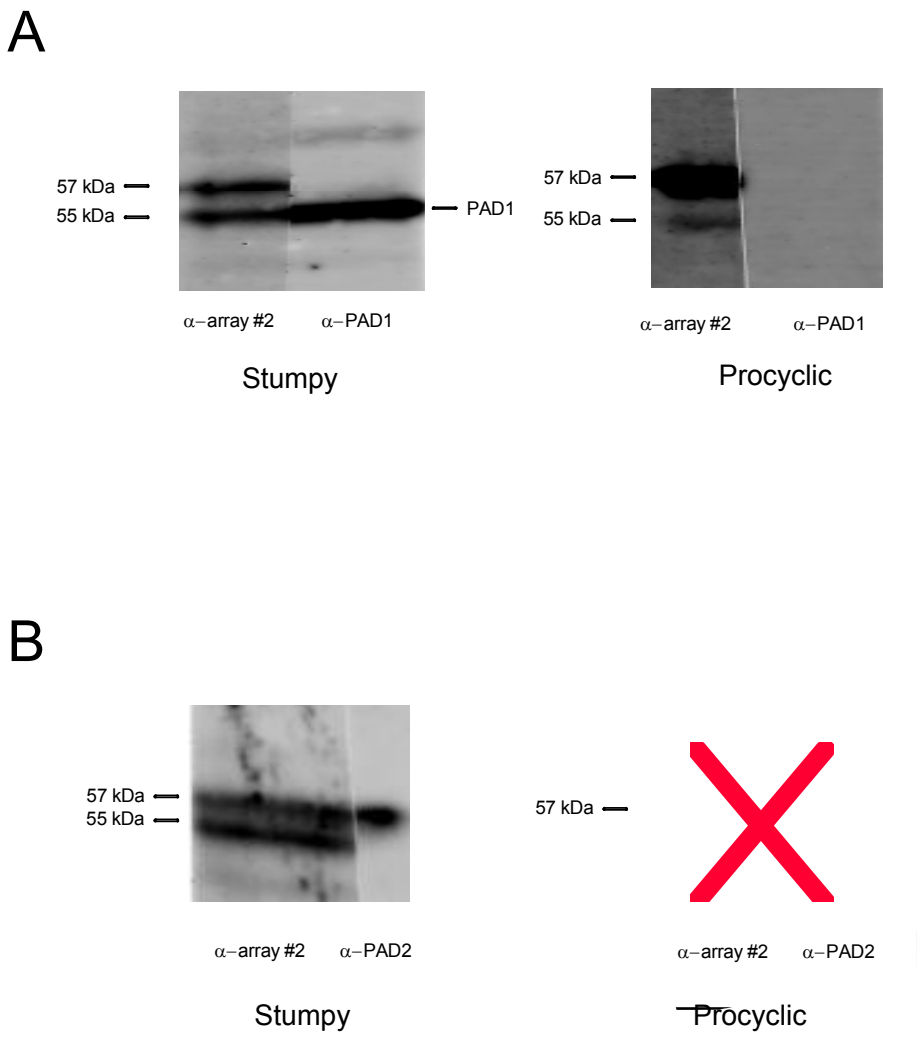


Figure 5.6-3 Identification of bands detected by the α-PAD array generic antibody. **C**ell lysate protein from different lifecycle stages was resolved by SDS-PAGE and transferred onto PVDF membrane. Post transfer, the protein was visualised by transillumination and the lanes were cut longitudinally, such that each half of the lane contained resolved protein. One half of each lane was stained with α-array #2, and the other half was stained with either (A) α-PAD1 or (B) α-PAD2 prior to staining with an α-rabbit secondary antibody. The lanes were subsequently aligned and visualised using the LI-COR odyssey system to determine which of the bands detected by the α-array antibody corresponded to either PAD1 or PAD2. The data show that the 55 kDa band in the stumpy form lane was PAD1, whereas the 57 kDa band in both lifecycle stages was PAD2.

5.7 Expression of PAD1 during synchronous differentiation to procyclic forms

Investigating the expression of PAD1 during differentiation may reveal a mechanism for PAD1 regulation. To this end, stumpy form AnTat 1.1 cells were purified from a late stage mouse infection, and differentiated by supplementing HMI9 media with 6 mM *cis*-aconitate and incubating at 27°C. Differentiating cells were analysed for PAD1 expression over a 24 hour timecourse by Western blot on cell lysate protein using the α-PAD1 antibody. PAD1 protein was quantified using the LI-COR Odyssey system.

Interestingly, Figure 5.7-1 shows that there appeared to be a slight (approximately 30%) upregulation of PAD1 protein in the first 2 hours of differentiation; this may be due to slender and intermediate cells differentiating to the stumpy form in response to differentiation conditions, or may be due to substrate mediated upregulation of PAD1, as has been observed with numerous other transporters (Diamond *et al.* 1984)) (Cheeseman and Harley (1991)) (Walker *et al.* (1998)) (Hu and Quick (2008)). Alternatively, this may simply represent some variation between the prepared samples in this experiment. However, at 24 hours there is still approximately 70% of the original PAD1 protein detected, suggesting that there is no specific mechanism for rapid clearance of PAD1 from the cell. Instead PAD1 may simply be removed from the cell through dilution once the cell has re-entered the cell cycle and undergone cytokinesis. No PAD1 protein was detected in established procyclic forms, in accordance with previous observations (see Figure 5.4-1).

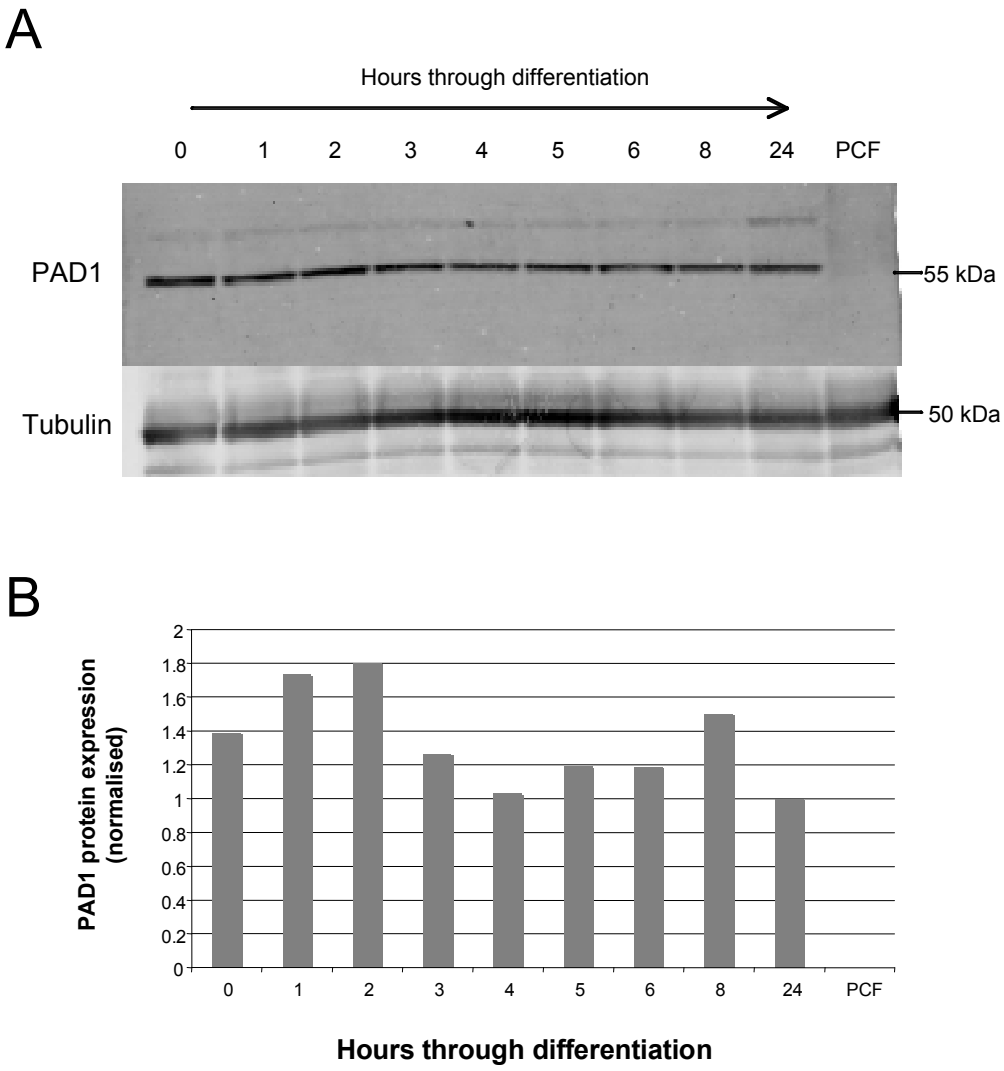


Figure 5.7-1 Expression of PAD1 through differentiation to the procyclic form. Stumpy form cells were induced to differentiate by incubation with 6 mM *cis*-aconitate at 27°C. Cell lysate was prepared over a 24 hour timecourse. (A) PAD1 expression was assayed by Western blot with the α-PAD1 antibody, using an α-tubulin antibody as a loading control. (B) The LI-COR Odyssey infrared quantification system was used to determine expression of PAD1 relative to tubulin throughout differentiation.

5.8 Expression of PAD2 through differentiation

It was of interest to determine how quickly PAD2 is upregulated during differentiation from the stumpy form to the procyclic form. Therefore, differentiating stumpy form cells were analysed for PAD2 expression by Western blot using the α-PAD2 antibody.

Figure 5.8-1 shows that PAD2 had relatively low expression in stumpy forms, but increased 2 fold after 1 hour. Expression fluctuated between 2 to 5 hours prior to increasing to maximal level of approximately 17 fold the expression found in stumpy forms at 24 hours. Expression in established procyclic forms was approximately 1/3 of that expressed at 24 hours post *cis*-aconitate addition, but was still approximately 6 fold more highly expressed than in stumpy forms. A 54 kDa band was detected that increased slightly in intensity; this probably represents either a degradation product or a differently resolving form of PAD2. Similar kinetics were detected using the α-PAD2 #1 antibody (data not shown). See Section 5.9 for further discussion.

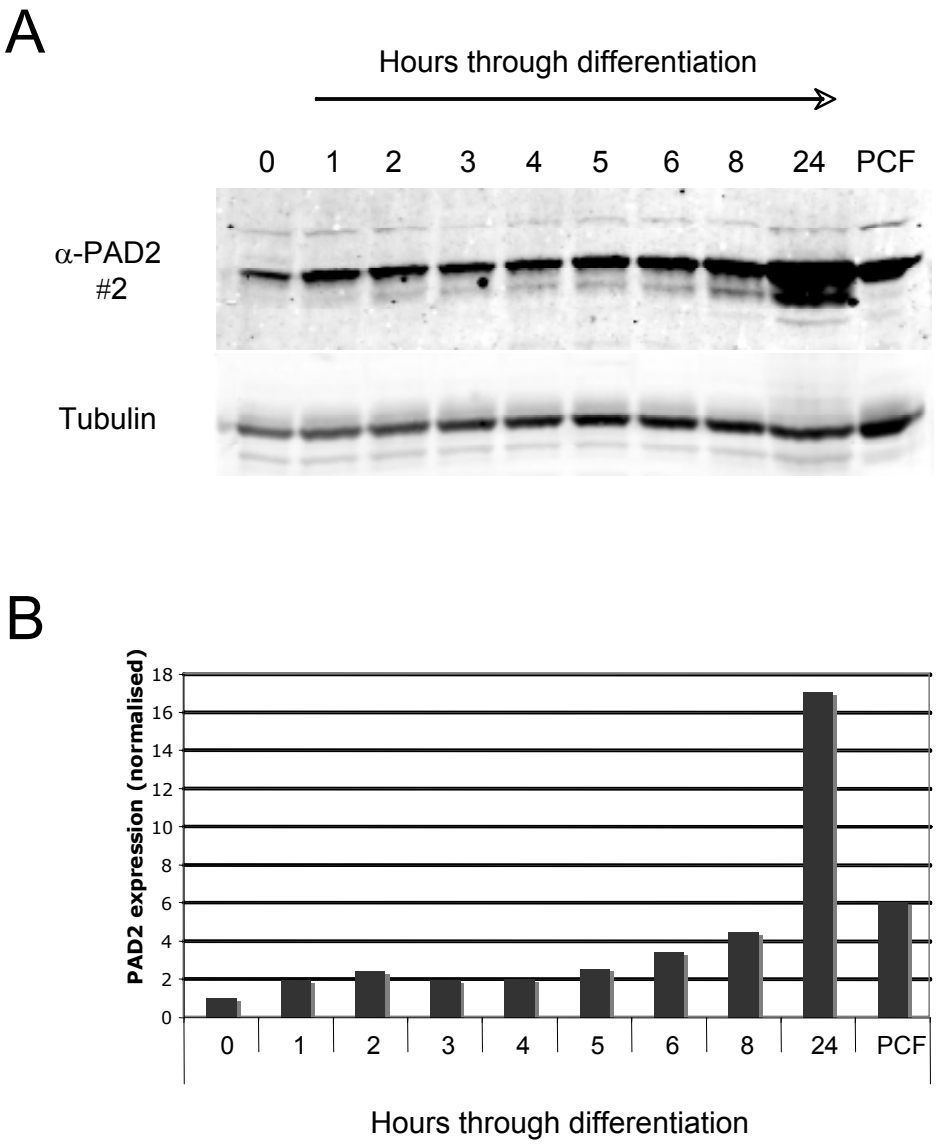


Figure 5.8-1 Expression of PAD2 through differentiation. Stumpy form cells were differentiation with *cis*-aconitate and protein cell lysate was prepared over a 24 hours timecourse. (A) Expression of PAD2 through differentiation was examined by Western blot analysis. An established procyclic form sample was also included. (B) PAD2 expression was quantified using the LI-COR odyssey system, and the data was processed in excel. Expression is normalised to a tubulin loading control, and expressed relative to the stumpy form lane.

5.9 Expression of entire PAD family of proteins through differentiation

PAD1 and PAD2 were shown to have distinct expression profiles both in the different lifecycle stages, and throughout differentiation to the procyclic form. It was therefore of interest to determine the expression of all of the PAD proteins throughout differentiation in order to (A) confirm the expression of PAD1 and PAD2 and (B) to determine whether any other PAD proteins were transiently expressed through differentiation. Additionally, if the 55 kDa band consisted of only PAD1, and the 57 kDa band consisted of only PAD2, then one would predict that the 57 kDa band would show the same expression profile as PAD2, and the 55 kDa band would show the same profile as PAD1. In contrast, if PAD proteins other than PAD1 and PAD2 (such as PAD5/7) contributed to either of these bands, and had a different expression profile, then one would predict that these bands would show a different profile to PAD1 and PAD2 during the differentiation time course.

Therefore, lysate from stumpy cells undergoing differentiation (see Section 5.7) was analysed by Western analysis using the α-array #2 antibody and the LI-COR Odyssey infrared imaging system in order to visualise and quantify PAD protein expression through differentiation.

Figure 5.9-1 shows the 55 kDa band precisely reflected the expression of PAD1 through differentiation, with an initial 30% rise in expression in the first 2 hours, followed by a drop at 5 hours prior to a slight increase at 8 hours. Although not detected at 24 hours, this was probably a result of the strong 57 kDa band masking detection of this protein. This provided further evidence that the 55 kDa band was PAD1; although PAD5 and 7 may share the same expression profile as PAD1, it is unlikely that they would resolve so closely to PAD1 as to be indistinguishable.

The 57 kDa band was consistent with the expression profile of PAD2 (compare Figure 5.9-1 and Figure 5.8-1) and showed an initial 2 to 3 fold increase after 1 hour. Similar to the expression profile determined using the α-PAD2 antibody, there was a slight fluctuation in signal between 2 to 5 hours, and maximal expression at 24 hours.

Established procyclic forms exhibit expression of the 57 kDa band at approximately 6 fold the expression of stumpy forms, and 1/6 to 1/7 that at 24 hours.

As has been stated, PAD2 expression was approximately 6 to 7 fold higher at 24 hours through differentiation than is detected in established procyclic forms. This has the interesting implication that this band not only represents a very early differentiation enriched protein, but that it is also much more highly expressed in an early procyclic form than a late procyclic form. It would be interesting to determine whether this protein responds to glycerol in a similar way to GPEET procyclin; GPEET procyclin expression is maintained in the presence of glycerol in the early procyclic form but is subsequently switched off (Vassella *et al.* (2000)). Note that continuously cultured procyclic forms lose repressional control over GPEET, and re-express this protein.

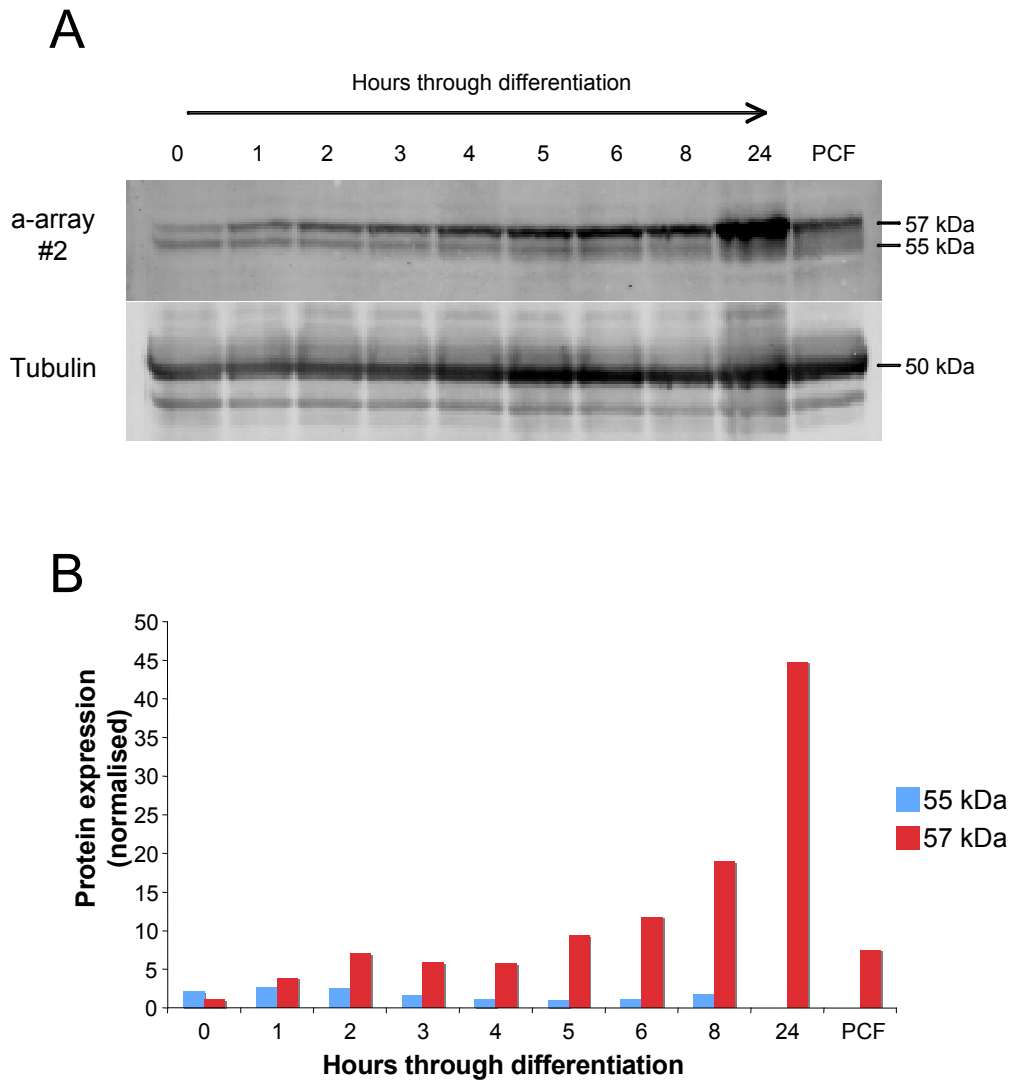


Figure 5.9-1 The expression of PAD proteins through differentiation. Stumpy form cells were supplemented with 6 mM *cis*-aconitate and cultured at 27°C for 24 hours. Protein cell lysate was prepared through this timecourse and analysed by Western blot using the α-array #2 antibody. (A) A Western blot to show expression of PAD proteins through differentiation (B) The signal from the Western blot was quantified and graphed relative to α-tubulin.

5.10 Expression of PAD1 is associated with expression of EP procyclin during differentiation of pleomorphic cells

There is an active debate in the trypanosome field about the capacity of slender cells to differentiate to procyclic forms. Serially passaged monomorphic cell lines have been selected such that stumpy forms are not present within the population. Despite this, monomorphic cells can differentiate to the procyclic form, albeit at a slower rate than a homogenous stumpy population. This has led to the proposal that true slender cells are competent for differentiation (Bass and Wang (1991)) and that initiation of differentiation is cell cycle stage independent (Li *et al.* (2003)). The alternative view is that stumpy cells are essential for lifecycle progression in pleomorphic cells, and that cell cycle arrest in G1/0 is required for differentiation (Matthews and Gull (1994)) (Tasker *et al.* (2000)) (Matthews *et al.* (2004)) (see Section 1.6 for further discussion).

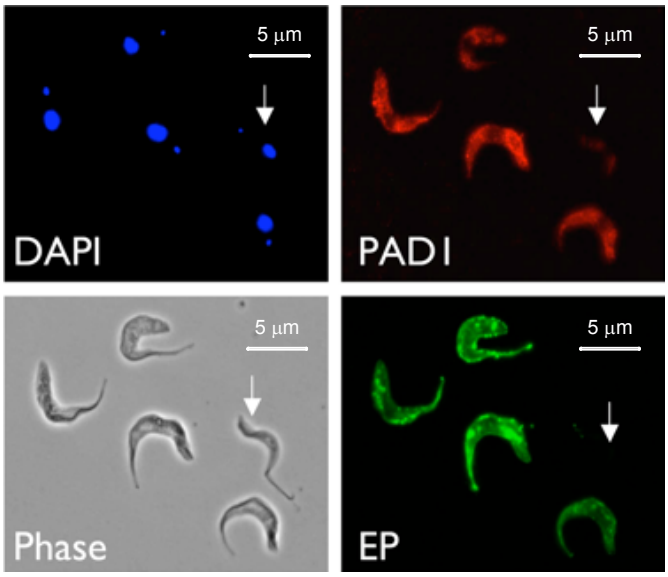
As a stumpy specific protein, PAD1 was used to investigate these conflicting hypotheses. Therefore, a pleomorphic population of cells containing stumpy, intermediate and slender form cells was purified from a mouse infection and induced to differentiate by addition of 6 mM *cis*-aconitate and incubation at 27°C. At 6 hours post *cis*-aconitate addition, the cells were fixed in methanol to preserve morphology and analysed by immunofluorescence to detect the expression of the procyclic marker EP procyclin, and the stumpy marker PAD1. The expectation was that if stumpy and intermediate cells are not required for differentiation then there would little or no correlation between the expression of PAD1 and EP procyclin. If, however, stumpy or intermediate form cells are an essential step in *T. brucei* lifecycle progression, then we should observe precise co-expression of PAD1 and EP procyclin on the cell surface.

Figure 5.10-1 shows that PAD1 expression correlated very closely with EP procyclin at 6 hours post *cis*-aconitate induction. More than 95% of the cells counted were either both EP positive and PAD1 positive, or EP negative and PAD1 negative. In

contrast, very few (less than 5%) were positive for one surface protein and not the other. Note that in the EP-ve/PAD1-ve population more than 80% of the cells exhibited 1 nucleus and 1 kinetoplast, (1K1N). Hence stumpy formation, rather than cell-cycle position alone, was required for EP procyclin expression.

Given that PAD1 has been established as a stumpy specific protein (see Figure 5.4-1), the strong correlation between expression of EP procyclin and PAD1 during differentiation strongly supports the hypothesis that the stumpy form is required for lifecycle progression in a natural population of *T. brucei*.

A



B

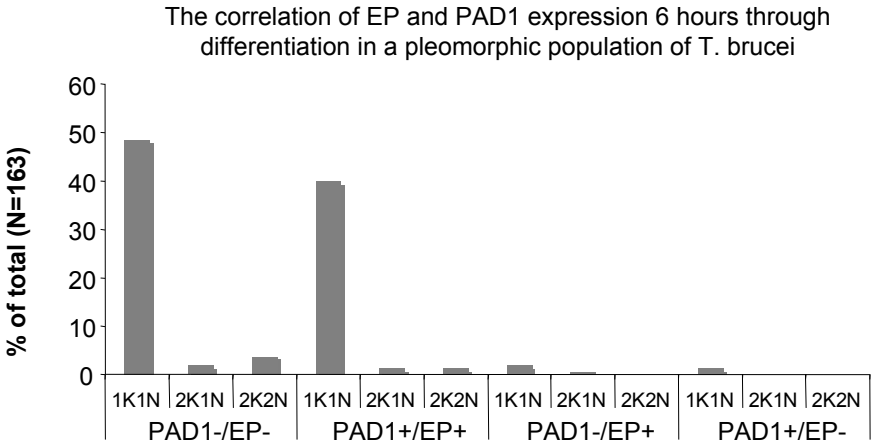


Figure 5.10-1 The correlation between PAD1 expression and EP expression during differentiation. The correlation of PAD1 and EP procyclin expression was determined in differentiating pleomorphic cells by immunofluorescence. (A) A representative field of view in which EP procyclin expression was precisely associated with PAD1 expression. The arrow indicates a cell that was not labelled with either marker. (B) Cells were scored as positive or negative for PAD1 and EP, and the position of the cell cycle was determined by counting the number of nuclei and kinetoplasts.

5.11 PAD expression in monomorphic slender cells

PAD proteins were not detected in monomorphic slender cells (see Figure 5.4-1). Thus, if the hypothesis that the PAD proteins are essential for reception of *cis*-aconitate is correct, the question arises: how are monomorphic cells able to differentiate in response to *cis*-aconitate? There are several examples of substrate induced upregulation of transporter molecules (Duan *et al.* (1999)) (Samuvel *et al.* (2005)) (Bernstein and Quick (1999)) (Munir *et al*. (2000)) and it may be that PAD1 is upregulated in the presence of *cis*-aconitate in monomorphic cells. This would have the interesting implication that *cis*-aconitate is not only able to induce differentiation to the procyclic form, but is also able to promote the slender to stumpy lifecycle progression. To test this possibility two strategies were attempted.

5.11.1 The affect of differentiation conditions on PAD1 protein expression in monomorphs

A healthy, exponentially growing monomorphic culture was subjected to four different treatments for 6 hours to identify which, if any, of the conditions for differentiation were able to upregulate PAD1 expression:

- 0 mM *cis*-aconitate; 37°C
- 6 mM *cis*-aconitate; 37°C
- 0 mM *cis*-aconitate; 27°C
- 6 mM *cis*-aconitate; 27°C

After 6 hours, cell lysates were prepared and Western blots were performed to detect any upregulation of PAD1, or any of the other PAD proteins.

Figure 5.11-1 shows that upregulation of PAD1 or the other PAD proteins was not detected in cells exposed to any of the above conditions. It may be that the high concentrations of *cis*-aconitate needed to stimulate differentiation of monomorphs forces signalling through other routes (such as a non-specific transporter).

Alternatively, it may be that PAD proteins were expressed at levels that are undetectable in this analysis. Nonetheless, no evidence of substrate-induced upregulation was detected.

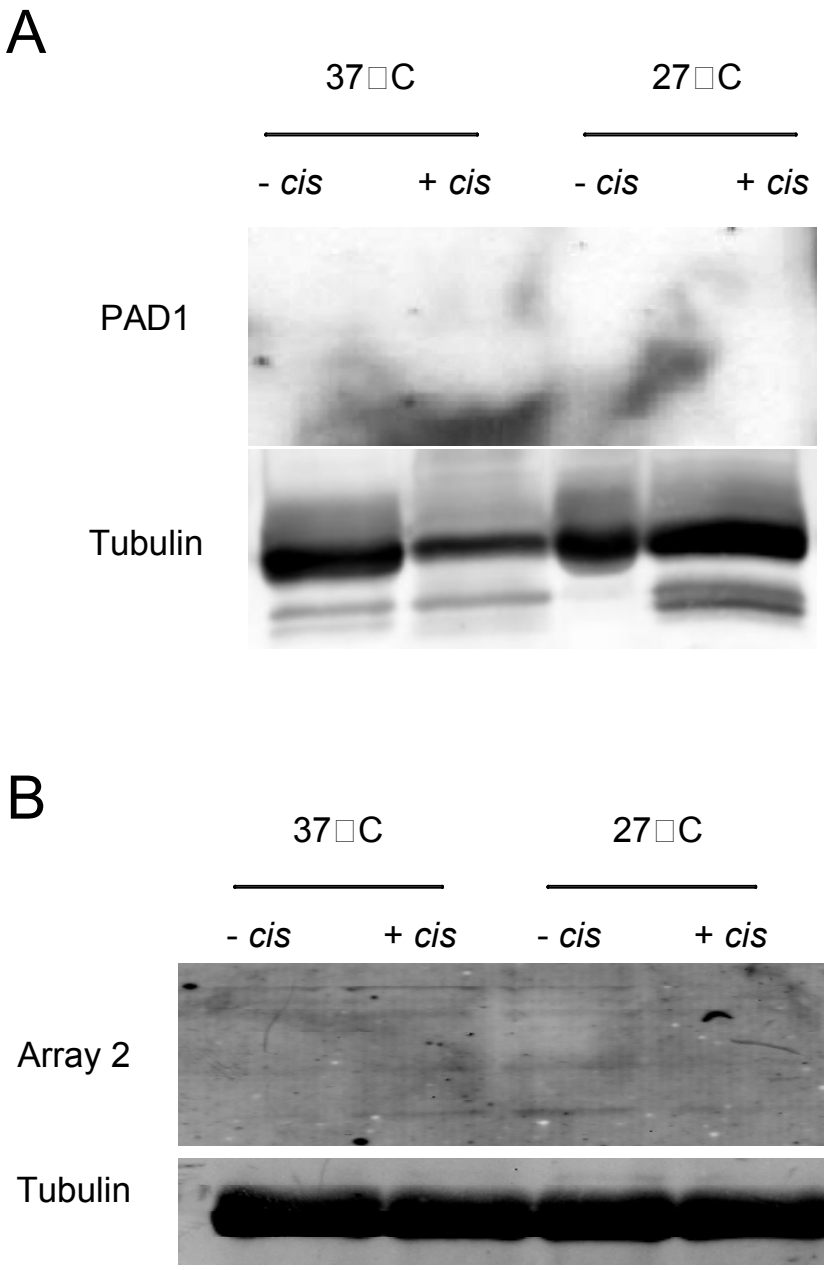


Figure 5.11-1 PAD expression under differentiation conditions. Lister 427 monomorphic slender cells were incubated with and without 6 mM *cis*-aconitate at 37°C and 27°C for 6 hours. Protein cell lysate was analysed by Western blot with (A) α-PAD1 and (B) α-array antibodies to determine whether PAD proteins are upregulated under differentiation conditions. No evidence of PAD protein upregulation was detected.

5.12 Investigating the kinetics of cold shock

Cold shock, as described by Engstler *et al.* (Engstler *et al.* (2004)), is induced by subjecting blood form trypanosomes to a drop in temperature from 37°C to 20°C. Under these conditions, stumpy forms express EP procyclin on their cell surface membrane and become hypersensitive to *cis*-aconitate. Slender cells also express EP procyclin in response to cold shock, but are unable to localise it to the cell surface and do not become hypersensitive to *cis*-aconitate. In order to investigate the affect of cold shock upon PAD1 protein expression it was necessary to first investigate the kinetics of cold shock. Specifically, the duration of temperature regulation needed to induce EP expression and its reversibility were assessed.

Thus, stumpy form cells were purified from a mouse infection and allowed to recover for 1 hour in HMI9 media at 37°C (see Materials and Methods for details). The culture was then divided such that half the cells were incubated at 20°C, and the other half were incubated at 37°C to control for potential inadvertent cold shock during cell purification. Thereafter, the cells were fixed in methanol at time intervals over 24 hours and the extent of cold shock assayed by immunofluorescence for EP procyclin expression. After 6 hours, a small proportion of cold shocked cells were re-incubated at 37°C to determine whether cold shock was reversible.

Figure 5.12-1 demonstrates that there appeared to be no detectable EP procyclin expression in response to cold shock within the first 2 hours, that by 3 hours a small proportion of cells were expressing EP procyclin, this rising to 27% by 5 hours. There was a small drop after 6 hours, whereupon procyclin appeared to remain stable at approximately 20% until 24 hours. Cells that were cold shocked for 6 hours, and then re-incubated at 37°C overnight, were able to reverse the cell surface expression of EP procyclin, such that only 3% of cells continued to express EP procyclin, compared to 20% of cells after 6 hours of cold shock.

Although expression of EP procyclin and hypersensitivity to *cis*-aconitate may not be coincident, we can conclude that cells are able to respond to cold shock within 4 hours, and that the cold shock affect is mostly reversible with respect to EP expression. This is useful technical information as it allowed design of experiments to investigate cold shock, and that, as long as due care is taken, we did not have be concerned with inadvertent cold shock of stumpy form cells during the purification process.

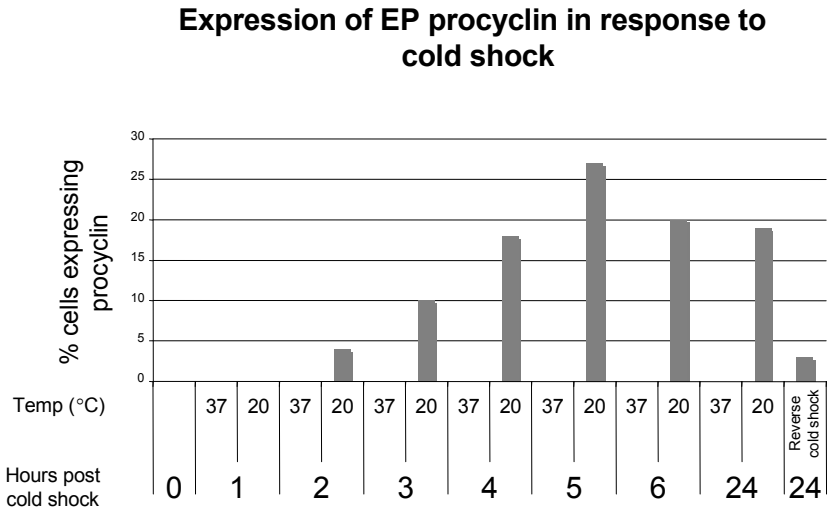


Figure 5.12-1 The kinetics of cold shock. AnTat 1.1 stumpy form cells were subjected to cold shock over a 24 hour timecourse. After 6 hours at 20°C, a small proportion of cells were re-incubated at 37°C overnight to determine whether the cold shock effect was reversible. The extent of cold shock was assayed by staining with EP-procyclin and analysed by immunofluorescence microscopy. 100 cells were counted *per* timepoint.

5.13 Cold shock does not cause increased levels of PAD1 mRNA

Given the putative role of PAD1 in *cis*-aconitate transport, it would not be surprising if PAD1 was implicated in a process that causes hypersensitivity to *cis*-aconitate in response to cold shock. This may occur through either PAD1 mRNA transcript stabilisation, increased translational efficiency of the PAD1 transcript, more efficient translocation of the protein to the cell surface membrane or a post-translational modification that increases *cis*-aconitate uptake.

To investigate the affect of cold shock upon expression of PAD1 mRNA, monomorphic slender cells, intermediate and stumpy form cells were incubated at 20°C for 18 hours; a control population was maintained at 37°C. RNA prepared from these cells was probed in a Northern blot using a DIG labelled RNA probe specific to PAD1 to detect any differences in PAD1 mRNA expression.

Figure 5.13-1 shows that there did not appear to be upregulation of PAD1 in response to cold shock in either monomorphic slender cells, intermediate or stumpy form cells. On the contrary, cold shocked intermediate and stumpy form cells expressed less PAD1 mRNA. This was unlikely to be a specific response to cold shock, but may be due to reduced efficiency of mRNA production at lower temperatures. As expected, monomorphic cells had almost undetectable levels of PAD1 mRNA until the blot was overexposed, and therefore probably showed no response to cold shock.

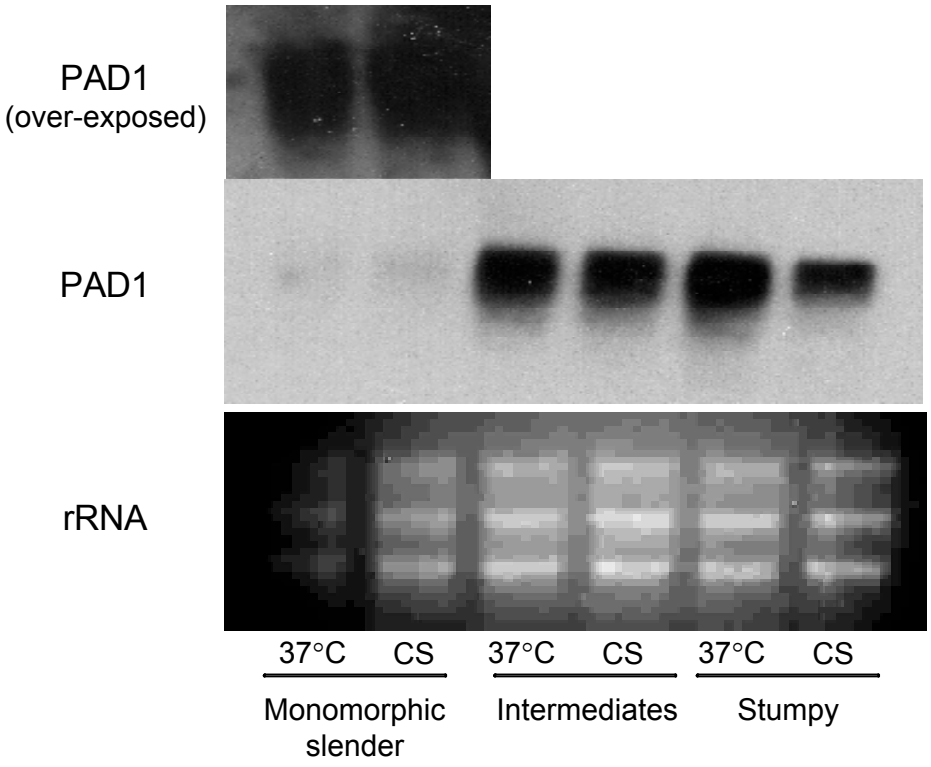


Figure 5.13-1 The affect of cold shock upon PAD1 mRNA levels in monomorphic slender, pleomorphic intermediate and stumpy form cells. Trypanosomes were incubated at 37° and 20° overnight and RNA was made for a Northern analysis of PAD1 mRNA levels. PAD1 RNA did not appear to up-regulated in response to cold shock, in contrast there may in fact be less transcript in the cold shocked intermediate and stumpy form cells, possibly due to reduced mRNA synthesis at lower temperatures.

5.14 There was a non-reproducible up-regulation of PAD1 protein in response to cold shock

In addition to mRNA, it was of interest to assess the impact of cold shock upon the levels of PAD1 protein in stumpy form parasites. Thus, stumpy form trypanosomes were purified from a mouse infection (described in Materials and Methods) and allowed to recover at 37°C for 1 hour in HMI9 media. Half the population of cells was subsequently incubated at 20°C whilst the rest remained at 37°C; after 18 hours samples were taken for Western, immunofluorescence and FACS analysis to detect expression of PAD1 protein using the PAD1 antibody.

Figure 5.14-1A shows by Western blot that there was a modest increase in PAD1 expression in response to cold shock; no such effect was detected in the monomorphic slender cell lysate. A flow cytometry analysis of PAD1 expression showed that a small proportion (approximately 5%) of stumpy form cells have increased expression of PAD1 protein (Figure 5.14-1B). Additionally, cells were analysed for PAD1 protein expression by immunofluorescence there was a clear upregulation of PAD1 in a small proportion of cells (approximately 5%); this upregulation was specific to the cold shocked cells and was not observed in cells cultured at 37°C. A detailed analysis of these cells showed that PAD1 high expressers were to be associated with an extreme stumpy morphology. This was judged by such attributes as a kinetoplast that was highly proximal to the nucleus, a 'comma' like shape, and no free flagellum at the anterior.

One might speculate that the cells that express PAD1 highly in response to cold shock are able to transport *cis*-aconitate into the cell at much lower extra-cellular concentrations, thus conferring the hypersensitivity to *cis*-aconitate reported by Engstler *et al.* (2004). Therefore, it was of interest to determine whether PAD1 high expression was associated with cold shock induced EP procyclin expression. For example, a correlation might reveal a mechanism for upregulation. To this end, duplicate slides of the original experiment were co-stained with α-EP procyclin and

α-PAD1, and counter stained with secondary antibodies that would distinguish between the mouse derived α-EP antibody, and the rabbit derived α-PAD1 antibody. However, no correlation was observed (data not shown).

Although three different types of data (Western analysis, FACS analysis and immunofluorescence) indicated that PAD1 was upregulated somewhat in response to cold shock in this experiment, subsequent repeats of this experiment did not show the same effect. A number of different variables were modified in an attempt to recapitulate the effect on PAD1 protein expression, and so establish the basis of its variability.

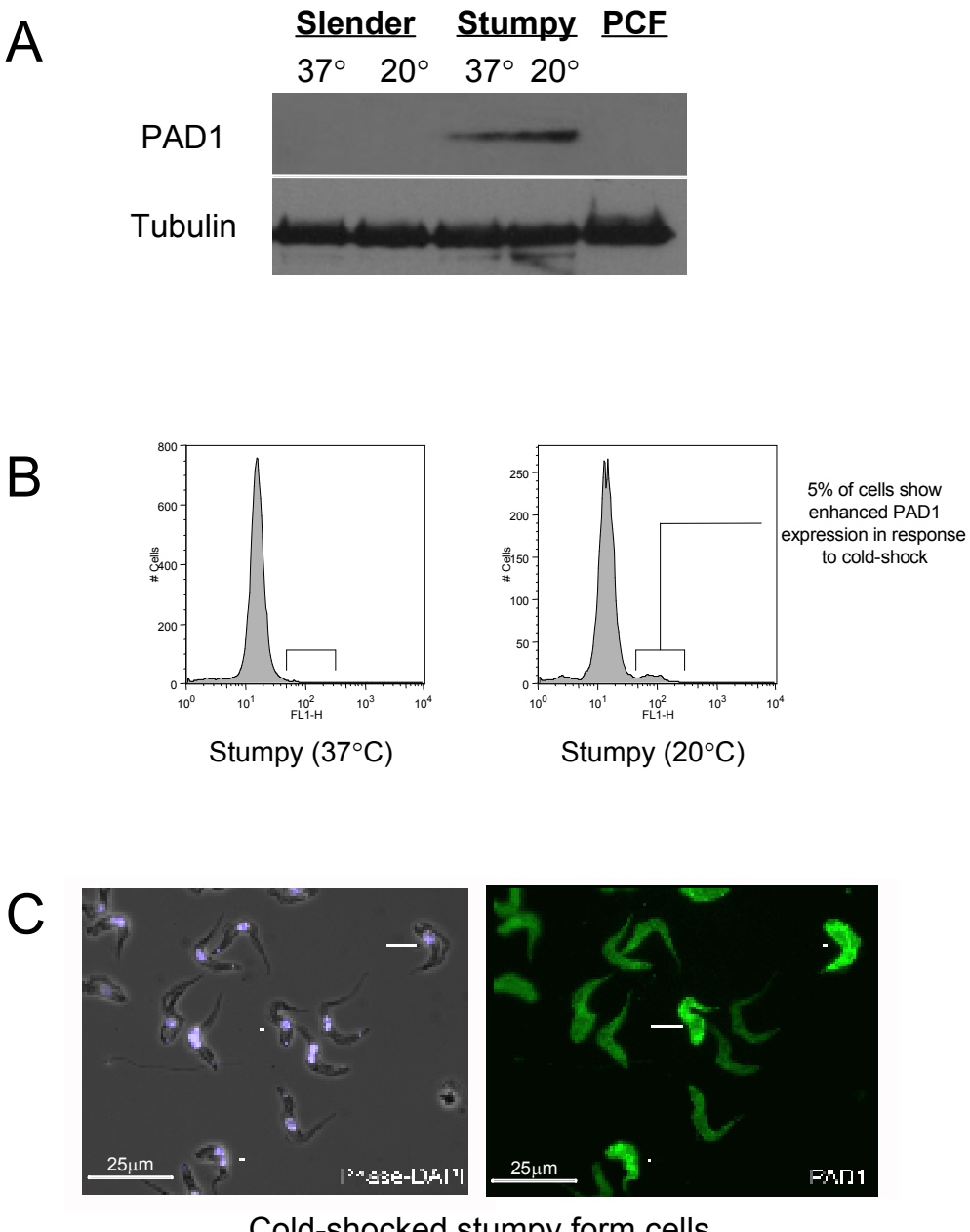


Figure 5.14-1 The effect of cold shock of stumpy form cells upon PAD1 expression. Stumpy form cells harvested from a mouse parasitaemia were subjected to cold shock for 24 hours and analysed for PAD1 expression by (A) Western analysis, (B) FACS analysis and (C) Immunofluorescence analysis. Monomorphic slender cells were also subjected to cold shock and included in the Western analysis with procyclic form cell lysate. In each a limited upregulation of PAD expression was observed, these being attributed to a small proportion of the population (panel C).

5.14.1 The effect upon PAD1 expression of varying the degree of parasitaemia and the temperature of cold shock

The cold shock of trypanosomes was originally reported to occur at 20°C, but it was possible that the experiment in Section 5.14 was subject to ambient temperature variations resulting in a lower temperature. Additionally, it was possible that the original experiment was performed using a relatively early stage parasitaemia and that intermediate cells are most responsive to cold shock. For these reasons both temperature and the degree of parasitaemia were tested for their affect upon cold shock and the expression of PAD1.

Therefore, two mice were treated with cyclophosphamide to depress their immune system and inoculated with AnTat 1.1 pleomorphic cells. Thereafter, the parasitaemia was monitored for stumpy cell formation by examining blood from tail snips for morphology. Each mouse was killed at the appropriate stage of parasitaemia such that one population had a mixed population of slender, stumpy and intermediate cells and the other had a homogenous stumpy population. Trypanosomes were purified using a DEAE cellulose anion exchange column and allowed to settle for 1 hour in HMI9 media. Subsequently, the cells were incubated at 37°C, 20°C, 15°C or 10°C in flasks with non-vented lids. After 18 hours, cell lysate was prepared for Western analysis, and cells were fixed with methanol for immunofluorescence analysis.

Figure 5.14-2A shows that PAD1 was not upregulated in stumpy form cells in response to any of the cold shock conditions. Moreover, when this experiment was performed on intermediate cells (see Figure 5.14-2B) the same result was observed, and indeed, a decrease in PAD1 protein expression was observed in response to lower temperatures. Further, immunofluorescence analysis of PAD1 expression did not show the small proportion of cells that upregulate PAD1 protein (data not shown). Thus, neither the temperature or the stage of parasitaemia induced elevated PAD1 expression.

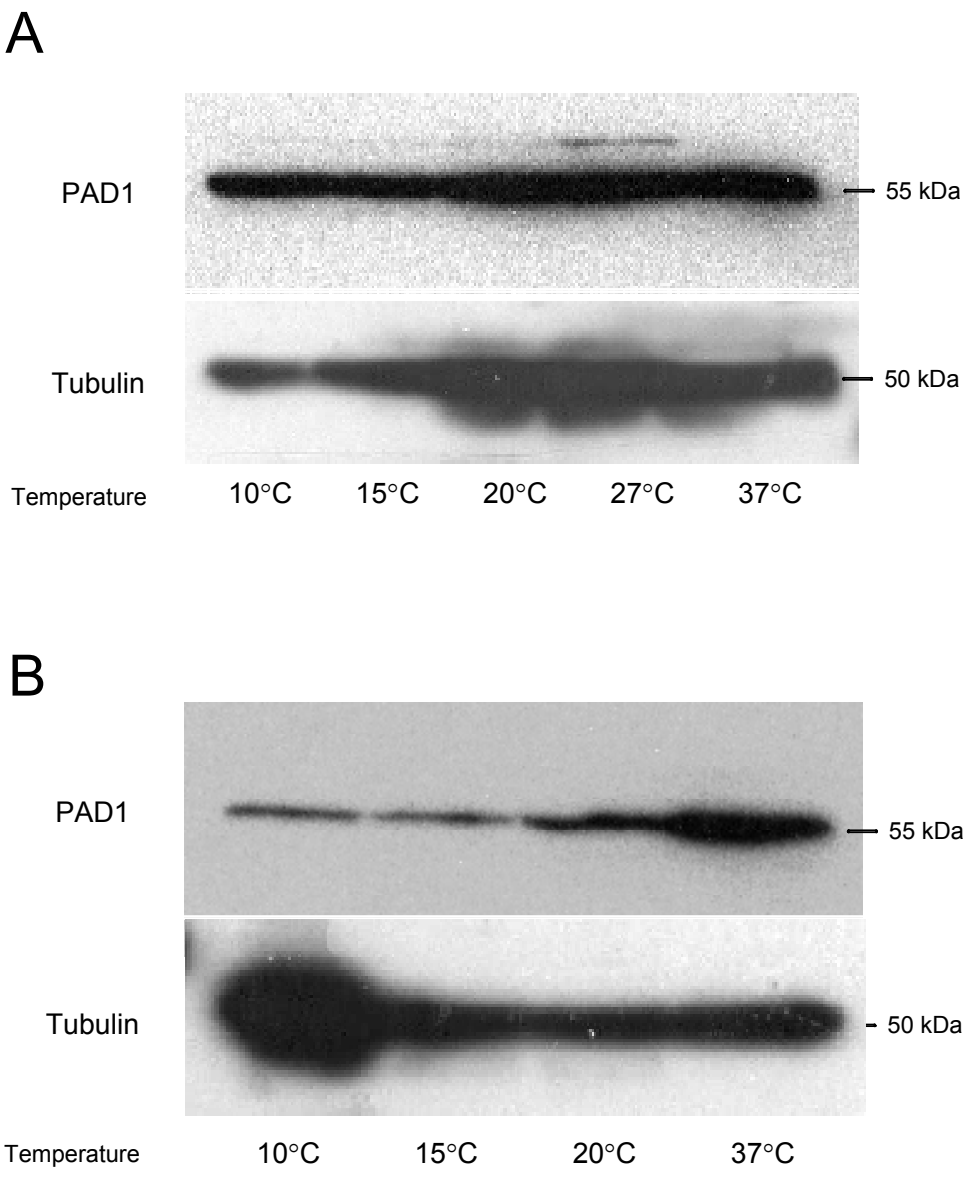


Figure 5.14-2 The affect of varying the temperature of cold shock and the degree of parasitaemia upon PAD1 expression. Stumpy form (A) and intermediate cells (B) were cold shocked for 18 hours at 10, 15 and 20 degrees Celsius (additionally, stumpy form cells were also exposed to 27 degree Celsius). Cell lysate was prepared for Western analysis to detect any change in PAD1 protein levels. The membranes were re-probed with anti-tubulin antibody as a loading control. PAD expression was not elevated at low temperature and, indeed, was somewhat decreased in panel B.

5.14.2 The affect of cold shock in combination with foetal calf serum (FCS) deprivation upon PAD1 expression

One might speculate that, in addition to a change in temperature, a change in the availability of nutrients might provide an environmental cue that the trypanosome has changed host, thus stimulating hypersensitivity to *cis*-aconitate through upregulation of PAD1. Additionally, there can be significant variations between bstches of foetal calf serum (FCS) that might explain the non-reproducibility of cold shock induced PAD1 upregulation. Therefore, it was decided to combine cold shock with FCS deprivation in an attempt to recapitulate the original PAD1 upregulation observed in Section 5.14.

Therefore, stumpy form trypanosomes were purified from a mouse infection (see Materials and Methods) and allowed to recover in HMI9 media containing 20% FCS at 37°C for 1 hour. Trypanosomes were then pelleted and re-suspended in media containing either 20% FCS, or no FCS at all. Half of the cells in each different condition were then incubated at 20°C in non-vented flasks, and the other half was incubated at 37 °C. After approximately 18 hours, cell lysate protein was prepared from each of the 4 cultures and analysed for PAD1 expression by Western blot. The LI-COR Odyssey system was used to quantify the signal relative to tubulin (see Materials and Methods).

Figure 5.14-3 shows a 55 kDa and a 59 kDa band when protein was reacted with the α-PAD1 antibody in a Western blot. The 59 kDa band is probably a differently resolving form of PAD1 that is sometimes detected in stumpy form lysate, but not in monomorphic slender cell lysate. There did not appear to be any convincing upregulation of PAD1 in response to FCS deprivation at either 37°C or 20°C; however a mild but clear (approximately 60%) upregulation of PAD1 was detected in response to cold shock conditions. This was surprising given the previous lack of success in recapitulating PAD1 upregulation in response to cold shock. This could reflect either a variation between protein samples unrelated to cold shock, the limitations of the LI-COR quantitation system, or an as yet unidentified variable.

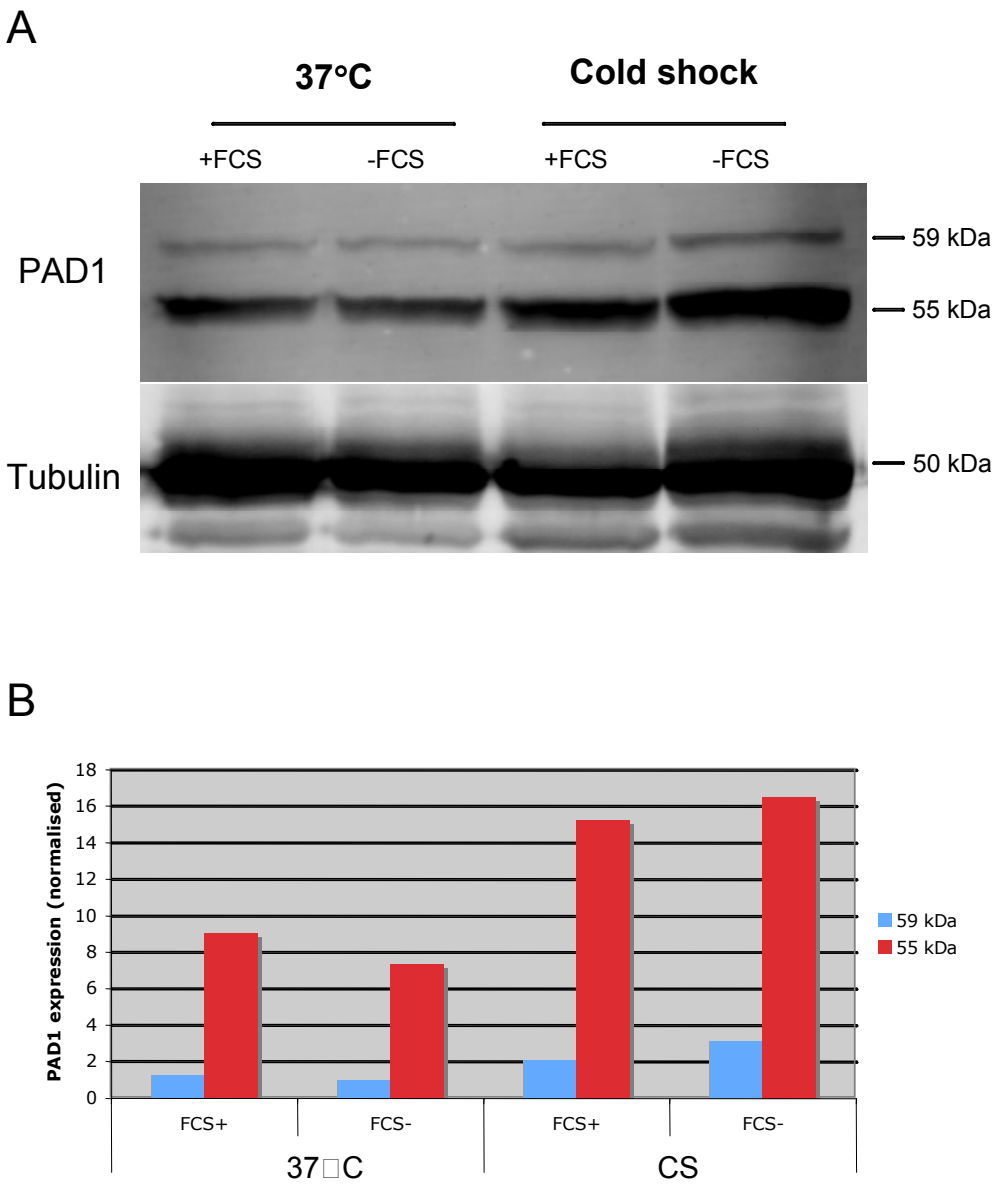


Figure 5.14-3 The effect of cold shock and FCS deprivation upon PAD1 protein expression. Stumpy form trypanosomes were harvested from mice and subjected to cold shock and/or foetal calf serum deprivation to assess the effect upon PAD1 protein expression. (A) PAD1 protein levels were assayed by Western blot analysis on protein cell lysate using the α-PAD1 #1 antibody. An α-tubulin antibody was included as a normalisation control. (B) Levels of PAD1 protein relative to tubulin were assayed using the LI-COR Odyssey infrared quantification system. PAD1 was upregulated in response to cold shock, but FCS deprivation had no significant affect upon PAD1 expression.

5.14.3 The effect of cold shock in combination with pH variation upon PAD1 expression

It has been reported that differentiation of stumpy form trypanosomes can be induced to differentiate to the procyclic form through exposure to mild acid conditions (pH 5.5) (Rolin *et al.* (1998)). Additionally, Rolin *et al.* also noted that mild acid treatment can promote the intermediate to stumpy transition and cause cell death in slender cells. From a physiological perspective, it is known that procyclic form trypanosomes grow in the tsetse in alkaline conditions (between pH 8 – 10) (Liniger *et al*. (2003)), but trypanosomes may be subject to transient pH fluctuations during uptake into the tsetse. Thus, pH may be an environmental cue that helps the trypanosome detect a change in host and stimulate hypersensitivity to differentiating factors such as *cis*-aconitate through upregulation of PAD proteins

To investigate this hypothesis, stumpy form cells were purified from a mouse parasitaemia as previously described and allowed to recover in HMI9, pH7.5 at 37°C for 1 hour. Cells were subsequently resuspended in HMI9 media at pH5.5, 7.5, 8 or 9. Half of the cells were then incubated at 20°C, and the remaining were incubated at 37°C for approximately 16 hours. Cell lysate was prepared for Western analysis, and cells were fixed in methanol for immunofluorescence analysis.

Figure 5.14-4 shows that there did not appear to be an upregulation of PAD1 in response to pH or cold shock under these conditions. The absence of increased expression in response to pH is surprising, given that mild acid conditions are reported to induce differentiation to the stumpy form, and would thus be predicted to cause upregulation of a stumpy specific protein. However, it is possible that PAD1 was already maximally expressed in these cells, and that mild acid treatment did not cause any further differentiation to the stumpy form. The reduction in PAD1 expression at pH5.5 may also reflect the induction of differentiation to procyclic forms, described by Rolin *et al.* (1998). Nonetheless, the lack of cold shock induced PAD1 upregulation underscores the lack of reproducibility of the original cold shock induced upregulation of PAD1 protein

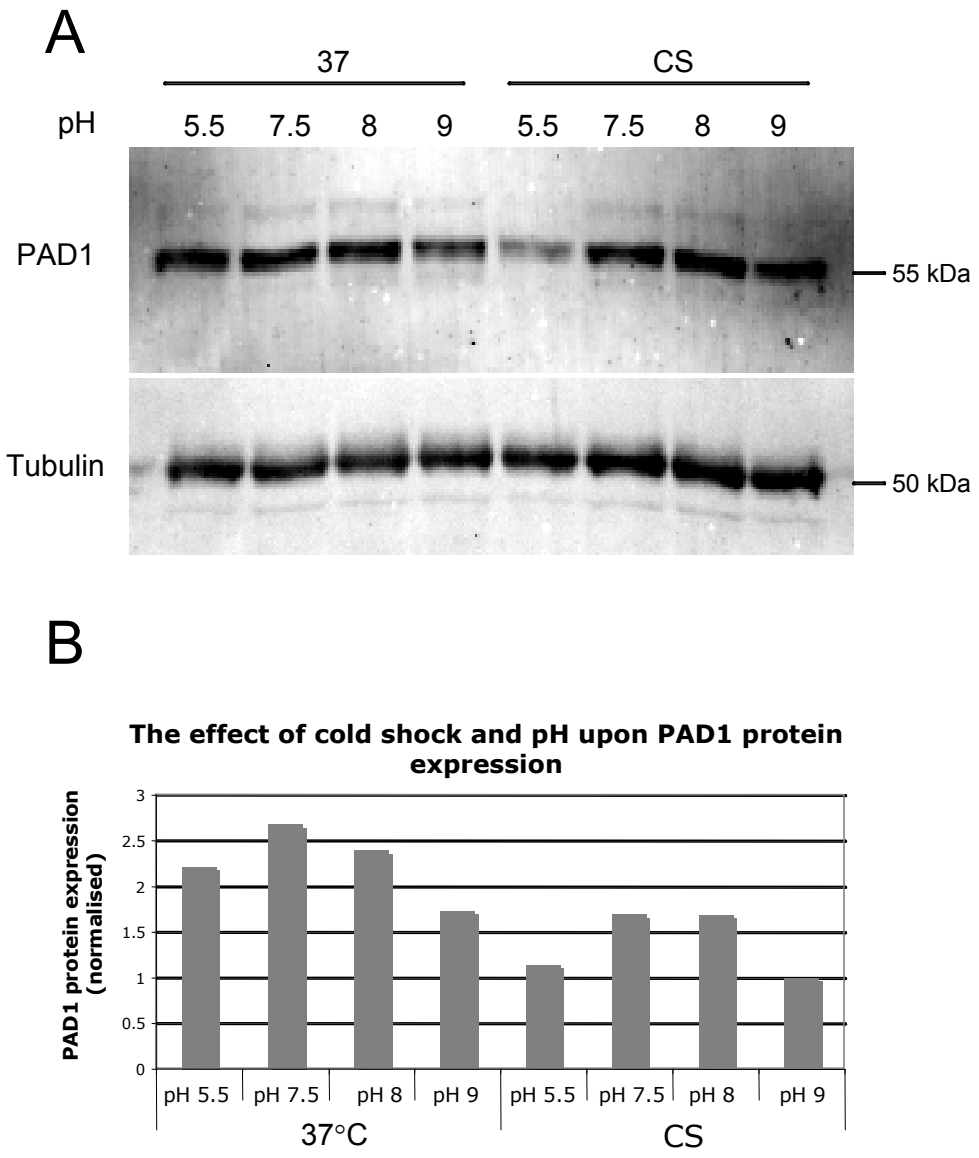


Figure 5.14-4 The affect of cold shock and pH upon PAD1 protein expression in stumpy form trypanosomes. Stumpy form trypanosomes were harvested from mice and subjected to cold shock in media of pH ranging from pH 5.5 to pH9. (A) A Western blot of cell lysate made after 24 hours of cold shock using the α-PAD1 antibody. Tubulin was included as a loading control. (B) The level of PAD1 expression relative to tubulin was assayed using the LI-COR Odyssey infrared quantitation system. Under these conditions, PAD1 expression is not upregulated in response to pH 5.5 or in response to cold shock. Cold shock appears to reduce PAD1 protein expression somewhat.

5.15 The effect of cold shock upon protein expression of other PAD members

Cold shock did not appear to cause upregulated expression of PAD1 protein, thus it was of interest to determine whether other PAD members are affected by cold shock. Cold shock was originally reported to induce expression of the procyclic form protein EP procyclin, and cause routing to the cell surface membrane of stumpy form cells only. This may be of particular relevant to PAD2, which has its maximal expression in procyclic forms.

5.15.1 The effect of cold shock upon the PAD family protein expression

The pleomorphic strains AnTat 1.1 and EATRO trypanosomes were used to inoculate mice in order to isolate stumpy form parasites. Stumpy form cells were purified from a mouse infection, resuspended in HMI9 media at approximately 5 x 10^6 cells ml^{-1}, and incubated overnight at either 37°C or 20°C. Protein cell lysate was prepared for Western analysis using the α-array #2 antibody (see materials and methods for details).

Figure 5.15-1 shows a mild, but clear and reproducible, upregulation of the 57 kDa band in response to cold shock. The extent of upregulation was variable between experiments, but was generally between 3 to 6 fold, with 1 experiment showing a 45 fold upregulation. As shown in Section 5.6.3, the 57 kDa band corresponds to PAD2, and subsequent data in this Chapter confirm this (see Section 5.16). Figure 5.15-1B also shows that the 55 kDa band is upregulated; but this is probably explained by the close resolution of the 57 kDa band interfering with the calculation of background signal, and does not appear to be supported when the exposure (Figure 5.15-1A) is examined.

Further experiment were performed in order to determine the effect of other variables, such as pH and foetal calf serum deprivation.

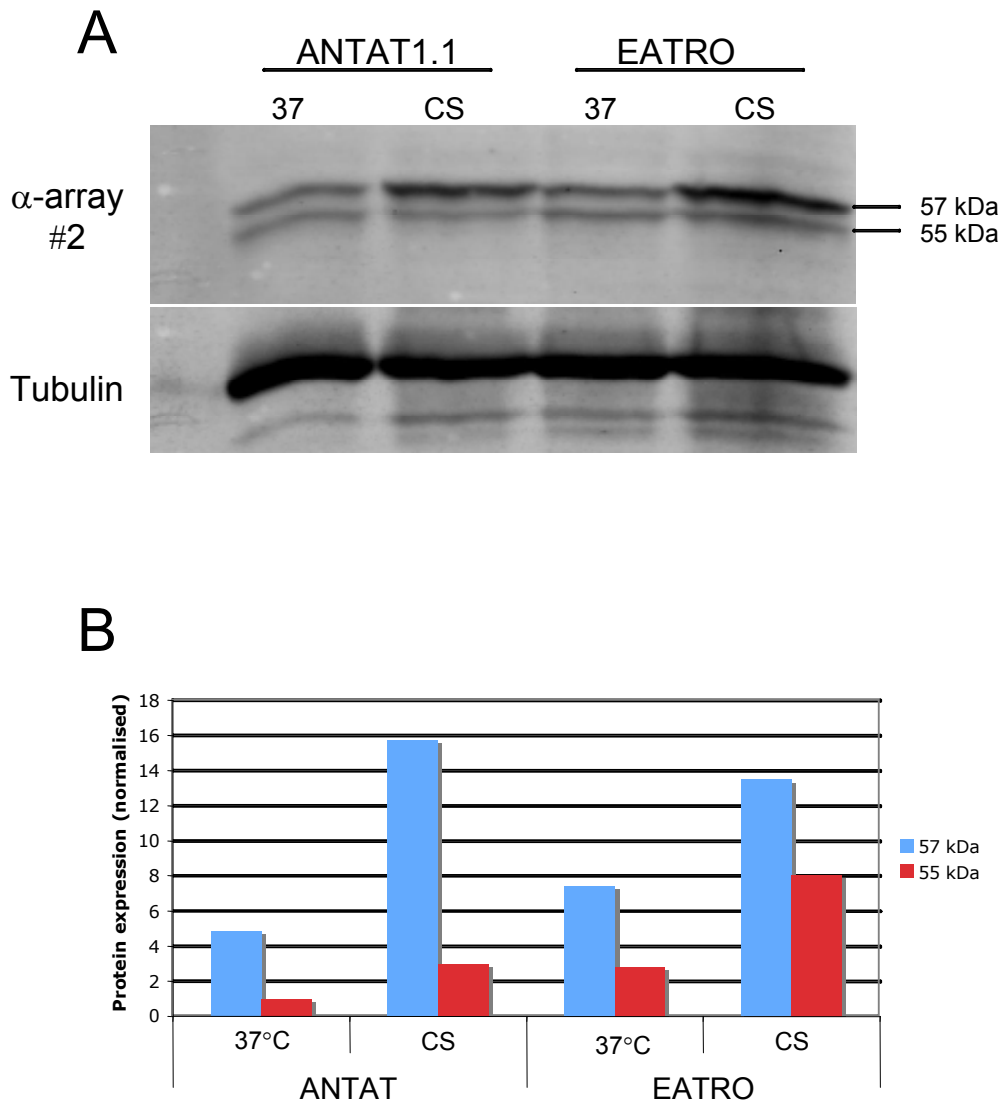


Figure 5.15-1 The effect of cold shock upon PAD family protein expression. AnTat 1.1 and EATRO stumpy form parasites were subjected to either 37°C or 20°C overnight. (A) Cell lysate was prepared, and analysed for PAD family protein expression by Western blot. (B) The resulting bands were quantified relative to the tubulin loading control using the LI-COR odyssey system, and the data was analysed. The 57 kDa band corresponding to PAD2 was 2 to 3 fold upregulated. This upregulation in response to cold shock was clear, although mild, and was consistently detected.

5.15.2 Western analysis PAD family proteins in response to cold shock and FCS deprivation

To determine the effect of foetal calf serum deprivation and cold shock upon PAD family protein expression, cell lysate from Section 5.14.2 was analysed using the α-array #2 antibody to detect all of the PAD proteins.

There were 3 bands visible in this analysis (see Figure 5.15-2). The 60 kDa band may be a form of PAD1 as this has been seen in Western blots using both the α-PAD1 antibody (see Figure 5.14-3) and did not change appreciably in response to FCS deprivation, cold shock or a combination of both. The 57 kDa band, corresponding to PAD2, did not show a convincing response to changes in the level of FCS, but did show a 3 – 4 fold upregulation in response to cold shock when cells were incubated in media containing 20% FCS, and approximately 2.5 fold upregulation when cells were incubated in media without FCS. The more limited cold shock-associated increase of this band in the absence of FCS was probably not significant and may have resulted from an impaired ability to synthesise new protein with limited nutrients. The 55 kDa band corresponding to PAD1 did not show a response to either cold shock or FCS deprivation, consistent with previous data (see Section 5.14.2)

In summary, there did not appear to be an upregulation of any of the PAD proteins in response to a nutrient deprived environment, but there did appear to be a reproducible 3 – 4 fold upregulation of the 57 kDa PAD protein in response to cold shock, consistent with previous data.

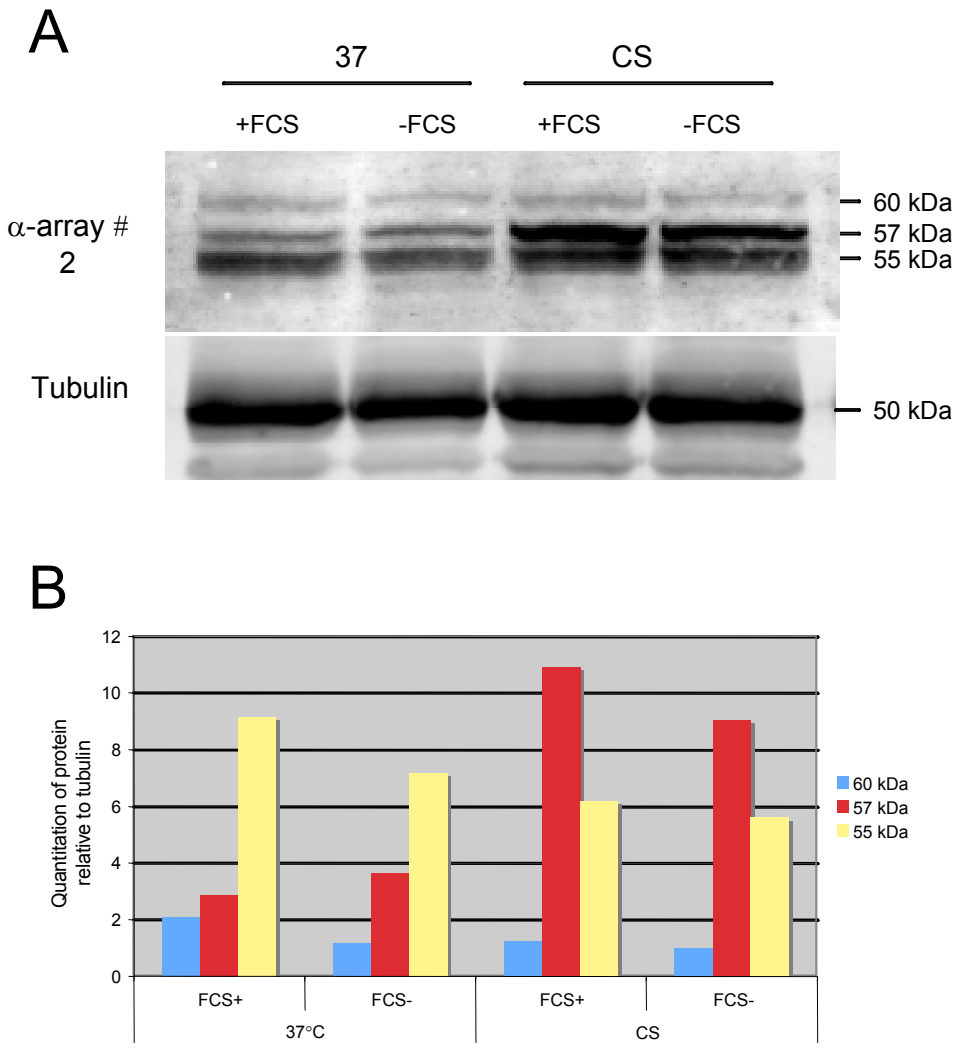


Figure 5.15-2 The affect of foetal calf serum deprivation and cold shock upon PAD array protein expression. Stumpy form trypanosomes were harvested from mice and subjected to cold shock and foetal calf serum deprivation to assess the affect upon PAD1 protein expression. (A) PAD protein levels were assayed by Western blot analysis on protein cell lysate using the α-PAD #2 antibody. An α-tubulin antibody was included as a control. (B) Levels of PAD family protein relative to tubulin were assayed using the LI-COR Odyssey infrared quantification system. The LI-COR data shows that the 57 kDa band responded to cold shock, but FCS deprivation had no significant effect upon expression of any of the PAD family proteins. The 60 kDa band is probably a form of PAD1 that resolves differently to the 55 kDa form of PAD1, and has been detected when using the α-PAD1 antibody.

5.15.3 The affect of pH and cold shock upon PAD array protein expression

Protein was analysed as for Section 5.14.3 using the α-PAD array #2 antibody. In this case, there appeared to be an up-regulation of the 57 kDa band by a factor of 45 (see Figure 5.15-3) in response to cold shock at pH7.5, and at least 3.5 fold in response to mild acid conditions at 37°C. The 55 kDa band does not show any convincing response to either cold shock or a change in pH.

The response of the 57 kDa band to pH5.5 does not necessarily represent pH regulation of this protein. Rather, as described in Section 1.5.3, cells are reported to initiate differentiation to procyclic forms at pH5.5 (Rolin *et al.* (1998)). Hence the apparent regulation at pH5.5 probably represents cell differentiation and concomitant elevated PAD2 expression seen in procyclic forms. Nonetheless, it is clear that under all conditions, cold shock induces significant upregulation of the 57 kDa band, corresponding to PAD2.

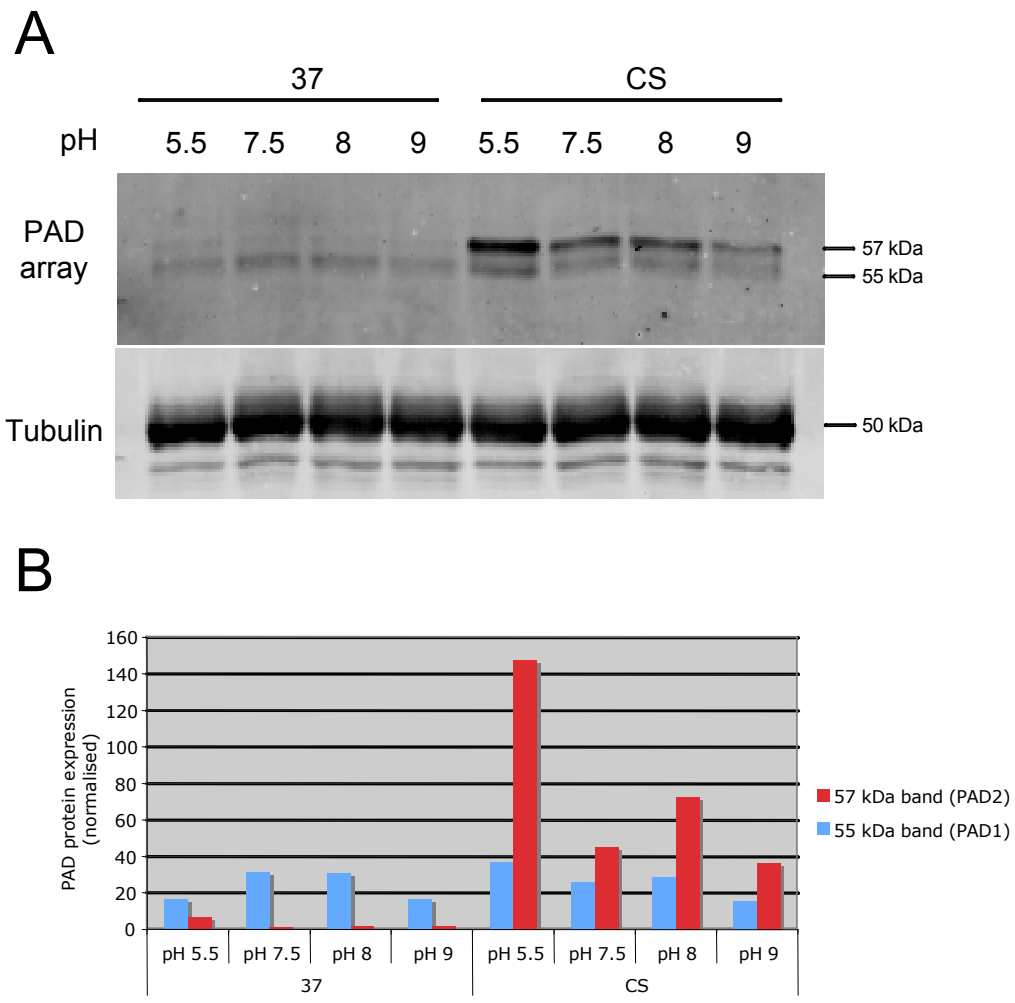


Figure 5.15-3 The affect of pH and cold shock upon PAD array protein expression. Stumpy form trypanosomes were harvested from mice and subjected to cold shock or 37°C in media of pH ranging from pH 5.5 to pH 9. (A) A Western analysis of cell lysate prepared after 24 hours of incubation at 20°C or 37°C using the α-array #2 antibody. Tubulin was included as a loading control. (B) The level of PAD array protein expression relative to tubulin was assayed using the LI-COR Odyssey infrared quantitation system. Consistent with earlier data, the 55 kDa band corresponding to PAD1 is not clearly upregulated in response to pH or cold shock. The 57 kDa band corresponding to PAD2 however, appears to show a strong (45 fold) upregulation in response to cold shock (consistent with previous data) at pH7.5, and a 3 to 7 fold increase in response to mild acid at 37°C.

5.16 The affect of cold shock upon PAD2 expression

Cold shock of stumpy form parasites causes a reproducible upregulation of the 57 kDa band detected using the α-array antibodies of between 3 to 45 fold (see Section 5.15). This band was shown to correspond to PAD2 (see Section 5.6.3). In order to verify whether PAD2 protein is indeed upregulated in response to cold shock, as previously described (see Section 5.14) stumpy form cells derived from three stocks were harvested from late stage mouse parasitaemias by DEAE cellulose anion exchange and subjected to either 37°C or 20°C (cold shock) overnight in HMI9 media. Cell lysate protein was prepared and analysed by Western blot for PAD2 expression using the α-PAD2 antibody.

Figure 5.16-1 shows that there was a clear and reproducible upregulation of the PAD2 band in response to cold shock in all 3 of the protein cell lysates analysed. Up-regulation was between 2 to 6 fold in the cold shocked samples, which is approximately what has been detected using the α-array antibodies for the 57 kDa band. Only the Western data for α-PAD2 #2 is shown, but α-PAD2 #1 shared a similar profile.

Thus, of the PAD family proteins, only PAD2 has been shown to display a reproducible upregulation in response to cold shock conditions. It would appear likely that PAD2 is responsible for the upregulation of the 57 kDa band detected by the α-array antibodies, although it cannot be completely excluded that another PAD protein or proteins also contribute to this signal.

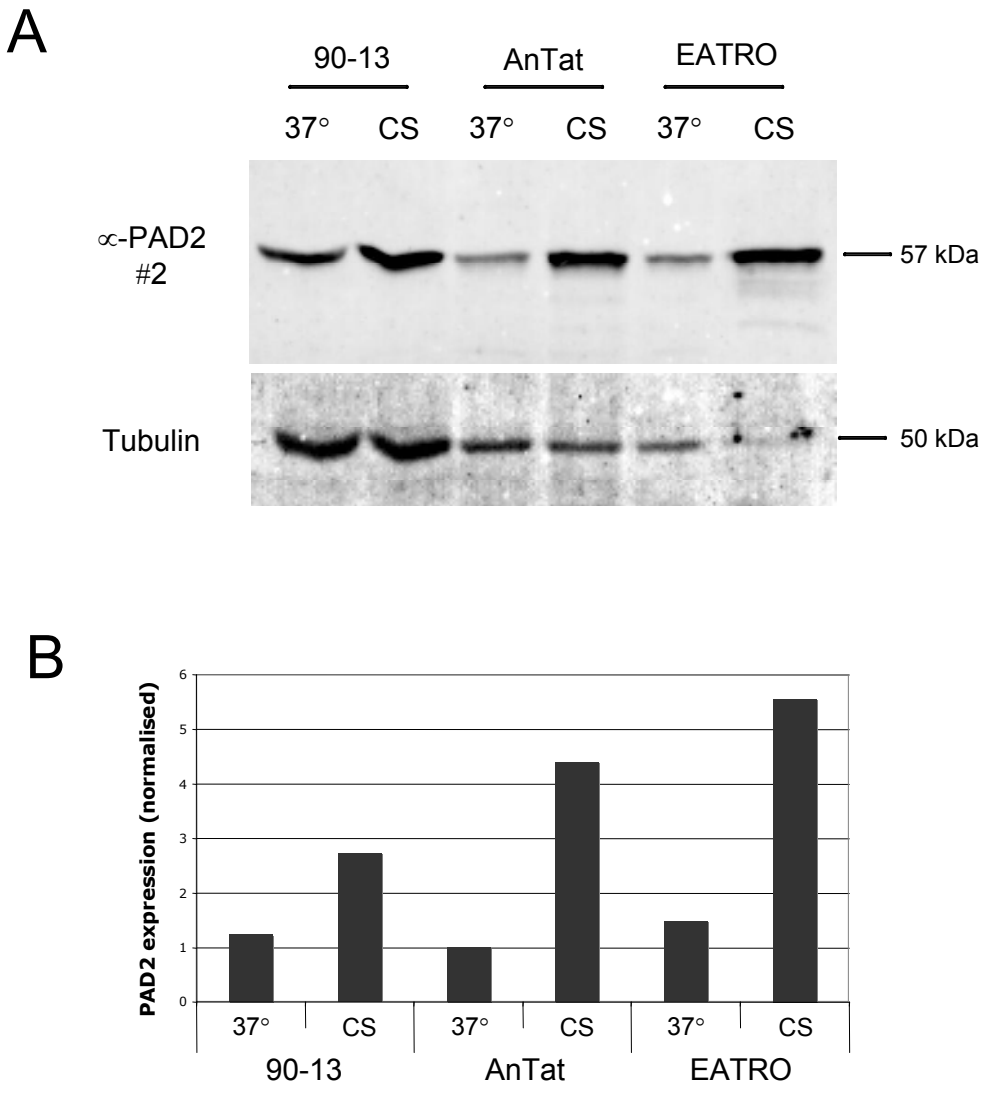


Figure 5.16-1 The affect of cold shock upon the expression of PAD2 protein. (A) AnTat 90.13, AnTat 1.1 and EATRO stumpy form parasites were harvested from a mouse parasitaemia and incubated at either 37°C or 20°C overnight in HMI9 media. Protein sample was prepared and analysed by SDS-PAGE and Western blot using the α-PAD2 antibodies and the LI-COR odyssey system. A tubulin loading control was included. (B) Analysis of the signal relative to tubulin using the LI-COR Odyssey system showed that there was a clear 2 to 5 fold increase in PAD2 protein expression in response to cold shock.

5.17 The affect of cold shock upon the localisation of PAD2 in stumpy form cells

Engstler *et al.* (2004) report stumpy form hypersensitivity in response to a temperature shift to 20°C. They further speculate that this could be due to a stumpy form specific protein that is only routed to the cell surface under cold shock conditions; at 37°C a cell surface trafficking checkpoint control mechanism prevents cell surface membrane localisation. The PAD2 protein upregulation detected by Western blot is relatively mild, and may not be sufficient to account for the dramatic increase in the sensitivity to *cis*-aconitate that has been reported. However, if PAD2 protein responds to cold shock in the way that the authors predict, then this would provide further evidence that PAD2 has a role in the reported hypersensitivity to *cis*-aconitate. To this end, experiments were designed to determine the effect, if any, of cold shock upon PAD2 localisation.

5.17.1 PAD2 localisation in stumpy forms at 37°C

Stumpy form cells were harvested from a late stage mouse parasitaemia and purified by DEAE cellulose anion exchange (see Materials and Methods for details), and recovered in HMI9 media. After 1 hour, half the culture was incubated at 37°C, and the remainder was incubated at 20°C, for 18 hours. Cells were subsequently analysed for PAD2 localisation by immunofluorescence microscopy using the PAD2 specific antibodies (see Materials and Methods). Images were captured using a confocal microscope, whereby the same settings were used to capture the 'secondary only' controls as were used when both primary and secondary antibodies were used.

Figure 5.17-1B shows only very faint, background signal when only the secondary antibody was used for staining. However, Figure 5.17-1A shows that when the α-PAD2 antibody is used as a primary antibody, the majority of cells (>80%) show a distinctive 'speckled' staining pattern. This staining pattern was not consistent with cell surface localisation as was shown for PAD1, and may reflect localisaion to internal vesicles. Note that a small proportion of cells (2 – 10%) exhibited the

classic cell surface pattern (see Figure 5.17-1A), with a staining around the periphery of the cell, suggesting that PAD2 is localised to cell surface in some stumpy form cells. This was repeated a number of times with similar results.

The staining pattern of PAD2 in stumpy forms incubated at 37°C suggests that rather than being exported onto the cell surface membrane, the majority of PAD2 is synthesised but remains within the endosomal system of the cell. Alternatively, it is possible that these speckles are cell surface bound, and that the staining pattern is due to cholesterol dependent microdomains, as been observed with some other membrane bound proteins (Prior *et al.* (2003)).

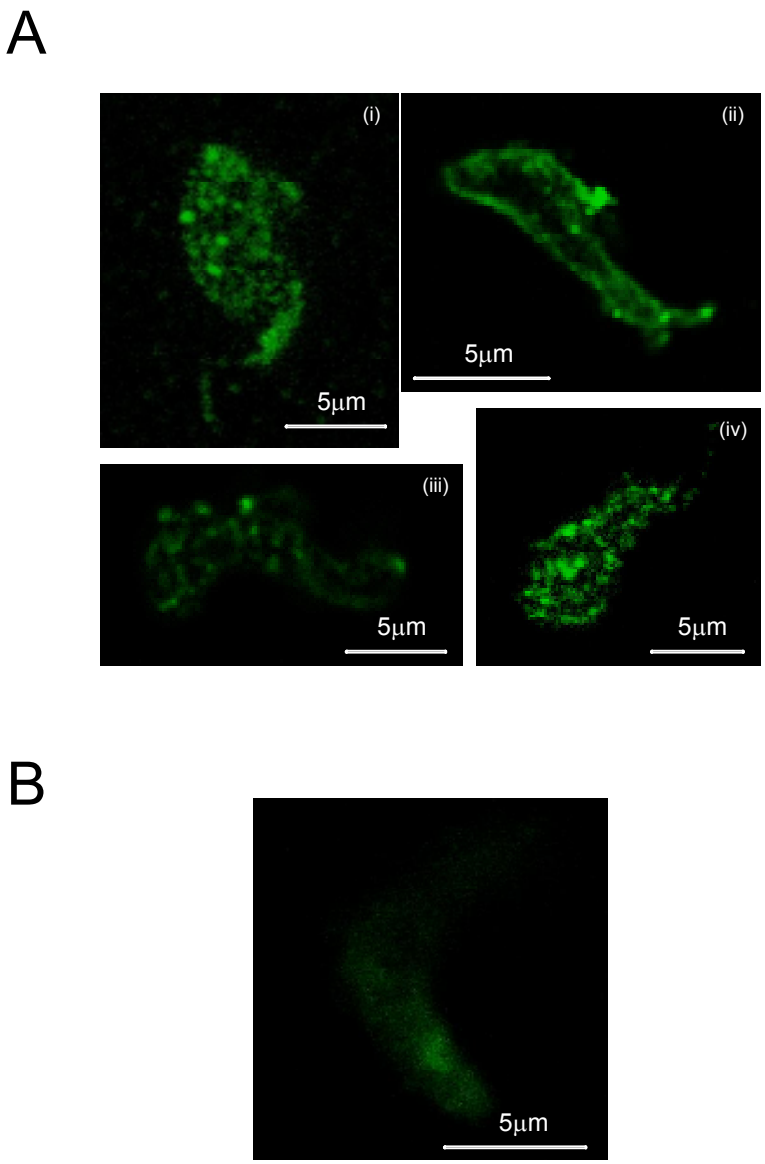


Figure 5.17-1 PAD2 localisation in stumpy forms at 37°C. (A) Stumpy form trypanosomes were stained with α-PAD2 and counter-stained with a FITC conjugated secondary antibody. Note that the image in (ii) shows peripheral staining, whereas the remaining cells demonstrated speckled staining. (B) Stumpy cells stained with the secondary antibody only do not show significant staining.

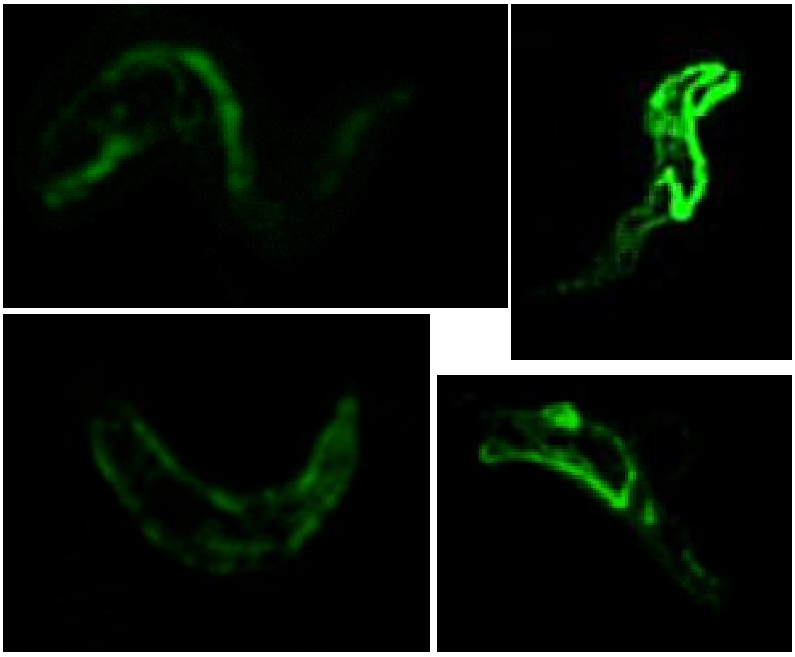
5.17.2 PAD2 localisation in stumpy forms at 20°C

To determine whether the cold shock induced upregulation of PAD2 effects PAD2 localisation, cold shocked stumpy form cells were stained with α-PAD2 and a FITC conjugated secondary antibody and analysed by confocal microscopy.

Figure 5.17-2A shows that there was a strong and consistent localisation of PAD2 to the cell surface membrane in cold shocked stumpy form cells. More than 90% of cells showed this pattern, with a small minority of cells displaying some speckles in addition to the cell surface pattern. Figure 5.17-2B shows that only a weak and diffuse signal is detected when the secondary only antibody is used with settings at maximum sensitivity.

Thus, it would appear that cold shock induces a change in localisation, in addition to an upregulation, of PAD2 protein. This type of system is not without precedent. As discussed (see Section 1.7.4.1 and Chapter 7), aquaporin is induced to re-localise from intracellular storage vesicles to the apical membrane of kidney epithelial cells in response to a signal cascade resulting from vasopressin binding (Kamsteeg *et al*. (2007)). This allows the expression of the transporter to the membrane with much faster kinetics than would be possible from upregulation of protein expression alone. The parallel with cold shock and hypersensitivity to *cis*-aconitate is clear. One might speculate that the trypanosome has evolved a rapid response mechanism, such that PAD2 is held within intracellular vesicles until cells are cold shocked, whereupon they release PAD2 onto the cell surface membrane and become hypersensitive to *cis*-aconitate.

A



B

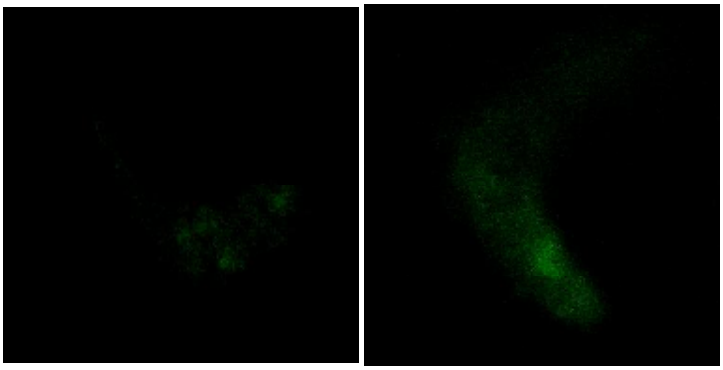


Figure 5.17-2 PAD2 localisation in cold shocked stumpy form cells. (A) Stumpy cells maintained at 20°C for 6 hours were analysed for PAD2 localisation by α-PAD2 immunofluorescence. PAD2 appears to localise to the cell surface membrane in cold shocked stumpy form cells. (B) Cold shocked cells were stained with only the FITC conjugated secondary antibody as a negative control. Only a weak, diffuse staining was detected.

5.18 Localisation of PAD2 in procyclic form cells

As discussed, PAD2 is predominantly expressed in procyclic form cells. To determine the localisation of PAD2 in this form, paraformaldehyde fixed cultured procyclic cells were stained with the α-PAD2 antibody and a FITC conjugated secondary antibody.

Figure 5.18-1A shows that PAD2 was localised to the cell surface membrane; staining was seen around the periphery of the cell, consistent with cell surface staining. Three representative cells are shown, and procyclic form cells universally exhibit this pattern, with no evidence of speckling or intracellular-like staining. Figure 5.18-1B shows a negative control cell that was stained with the secondary antibody only; a very weak pattern was detected that did not account for the signal obtained when the primary antibody was included. Thus it would appear that procyclic cells also express PAD2 at the cell surface membrane.

Thus, PAD2 appears to be localised to intracellular vesicles at 37°C. Upon cold shock, PAD2 appears to be re-localised to the cell surface membrane, where it stays as the cells differentiate to the procyclic form. It would be interesting to determine the expression and localisation of PAD2, and the other PAD proteins, in mesocyclic, epimastigote and metacyclic forms.

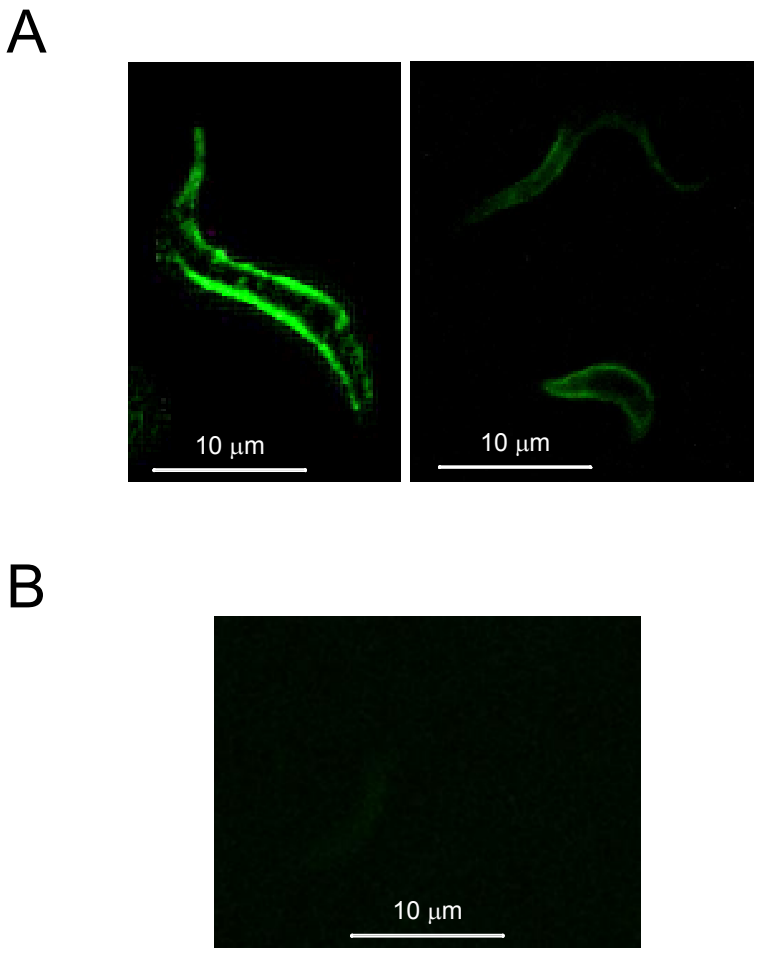


Figure 5.18-1 PAD2 in procyclic form cells. (A) Procyclic form cells were analysed for PAD2 localisation by staining fixed cells with the α-PAD2 antibody and a FITC conjugated secondary antibody. The cells universally expressed PAD2 on the cell surface membrane. (B) A secondary antibody only negative control is shown. The signal shown in (A) appears to be due solely to the presence of PAD2.

5.19 Conclusion

In summary, the PAD proteins are differentially expressed in different lifecycle stages, and in response to different environmental conditions. Additionally, conditions reported to induce hypersensitivity to *cis*-aconitate, cause increased expression and re-localisation of one member of this family.

The data presented here support the hypothesis that the PAD proteins promote differentiation by transporting citrate/ *cis*-aconitate into the cell. A functional analysis was therefore performed to test this hypothesis.

Chapter 6

An analysis of the PAD Family Function

6 An analysis of PAD family function

The most common route to determine the function of a newly identified protein is to perturb the level of its expression and assay for a phenotypic effect. In the case of the PAD family there were some inherent technical difficulties associated with this strategy. PAD proteins were shown not to be expressed in monomorphic slender cells (see Sections 5.4, 5.5 and 5.6), thus the conventional approach of RNAi and generating null mutants in this cell type were unlikely to yield a phenotype. However, many groups have encountered insurmountable difficulties in performing reverse genetic studies in stumpy form cells; pleomorphic cells are extremely difficult to culture, and are thus resistant to transgenic studies. In consequence, a significant amount of work was invested in genetically manipulating monomorphic cells to determine whether any differentiation phenotype was observed after PAD expression was perturbed. This was justified since the PAD genes were originally identified as being upregulated in a monomorphic cell line that is unresponsive to *cis*-aconitate. Hence, it was clear that there may be an association between PAD gene expression and differentiation, even in monomorphic cells.

6.1 The effect of ectopic expression of PAD1 upon differentiation

Ectopic expression of PAD1 was previously shown not to affect the differentiation of monomorphic slender cells in response to 6 mM *cis*-aconitate (see Section 4.5). However, the high extracellular concentrations of *cis*-aconitate used in these experiments meant that entry of *cis*-aconitate into the cell may not have been the limiting factor in stimulating differentiation . Also, the earlier experiment made use of a cell line that was ectopically expressing a Ty epitope tagged form of PAD1. This may have interfered with the protein's function or substrate interaction.

6.1.1 The effect of expression of an ectopic copy of PAD1 upon the sensitivity to *cis*-aconitate at 20°C and 37°C

Ectopic expression of PAD1 did not cause an increase in the number of trypanosomes that are able to respond to 6 mM *cis*-aconitate, thus differentiation capacity remained the same (see Section 4.5). However, it may be that cells expressing high levels of PAD1 on the cell surface membrane have increased efficiency of *cis*-aconitate transport, and are therefore able differentiate at lower concentrations of *cis*-aconitate. Also, Engstler *et al.* (Engstler *et al.* (2004)) showed that, when stumpy form cells are subjected to cold shock conditions of 20°C, they become hypersensitive to *cis*-aconitate, being able to respond to a 500 fold lower concentration of *cis*-aconitate. This effect is specific to stumpy form cells and was not exhibited in true slender cells, or monomorphic slender cells (Samuel Dean, unpublished observations). However, as a stumpy form specific protein, it may be that PAD1 expression is the determinant that enables stumpy form cells to become hypersensitive to *cis*-aconitate in response to cold shock.

Two possibilities were investigated in the experiment described below: (i) that upregulation of PAD1 is necessary and sufficient for the cold shock effect; thus by upregulating PAD1 we would predict that the trypanosomes would exhibit increased sensitivity to *cis*-aconitate at 37°C, and (ii) that upregulation of PAD1 is necessary, but not sufficient, for hypersensitivity to *cis*-aconitate; thus upregulation of PAD1 causes increased sensitivity to *cis*-aconitate at 20°C, but not at 37°C. The second hypothesis relies on a cold shock inducible factor being expressed in monomorphic cells, which may or may not be the case.

In order to investigate these possibilities, a cell line was generated using the expression vector pHD617 that had been modified to contain a puromycin resistance cassette in place of a hygromycin cassette (see Materials and Methods for details, and Appendix A for a plasmid map). The expression vector pHD617 is optimised for expression in blood form trypanosomes and was expected to give higher expression than pHD451 (Biebinger *et al.* (1997)). The cell line was engineered such that PAD1 could be inducibly expressed from the ribosomal spacer region of the genome when

the cells were cultured in the presence of tetracycline. These cells were cultured with and without tetracycline to induce expression of PAD1; parental trypanosomes containing only pHD449 were treated equivalently as a negative control. After 5 days of induction at 37°C, half of each of the four cultures was incubated overnight at 20°C so that the following conditions existed for both the PAD1-expressing cell line, and the parental control:

- + tetracycline: 37°C
- - tetracycline: 37°C
- + tetracycline: 20°C
- - tetracycline: 20°C

After the overnight incubation, each of the 8 cultures was induced to differentiate using different concentrations of *cis*-aconitate: 6 mM, 1 mM, 10 μM and no *cis*-aconitate. The 32 cultures were incubated at 27°C and samples were taken at 0, 24, 48 and 72 hours for FACS analysis to assay their progression through differentiation. Additionally, cell lysates were prepared prior to the addition of *cis*-aconitate for Western analysis to determine the relative levels of PAD1 expression.

Figure 6.1-1A shows that cells induced to express PAD1 respond slightly better than non-induced cells to 1 mM *cis*-aconitate at 37°C after 24 hours (22% procyclin positive, versus 11% respectively), but this was not observed at later time-points or at 10 μM *cis*-aconitate and was probably experimental variation unrelated to PAD1 expression. The cell line containing the ectopic copy of PAD1 differentiated at slightly higher efficiency than the untransfected cell line when exposed to 6 mM *cis*-aconitate at 37°C irrespective of the presence of tetracycline (82% differentiation after 48 hours, compared with approximately 62% for the untransfected), but this was not inducible and was also probably unrelated to the expression of PAD1.

When these cells were subjected to cold shock prior to addition of *cis*-aconitate, the cell lines transfected with ectopic PAD1 appeared to be inhibited in their ability to differentiate in response to 6 mM *cis*-aconitate when compared with the 427 control

(approximately 30% of cells have differentiated compared with the untransfected cell line). However, there was no inducible effect related to PAD1 expression and so the differences between the cell lines were probably unrelated to expression of the *trans*-gene. Moreover, neither of the cell lines were able to differentiate in response to less than 6 mM *cis*-aconitate, indicating that PAD1 expression did not confer enhanced sensitivity to *cis*-aconitate.

A Western blot using the α-PAD1 antibody was performed on protein samples to determine PAD1 expression prior to the addition of *cis*-aconitate. No PAD1 expression was detected in any of the monomorphic cell lysates from cells maintained at 37°C. This is despite the earlier inducible expression of PAD1-Ty (see Section 4.5). This may be due to a less efficient expression of PAD1 in this construct (despite its optimisation for blood form expression), or could be due to the α-PAD1 antibody being less sensitive than the BB2 antibody. Interestingly, PAD1 is detectable at low levels in the cold shocked cells. Expression is leaky, but analysis of PAD1 signal using the LI-COR Odyssey system shows that addition of tetracycline causes 3 fold increased expression. However, expression is far lower than in stumpy forms, whereby PAD1 is 10 fold more highly expressed than even the induced cells.

The detection of PAD1 in only the cold shocked cells may be due to an accumulation of protein unrelated to a specific biological mechanism. The vector pHD617 utilises the aldolase 3' UTR to confer mRNA stability, and has not been reported to show any effect of cold shock. It is possible that the upregulation is mediated at the post-translational level, possibly through molecular chaperones conferring enhanced folding and therefore translation. However, the most likely explanation for this effect is probably that cell division was greatly inhibited at 20°C (data not shown), thus expressed protein was not diluted as quickly through cell division.

Nonetheless, no convincing effect of PAD1 ectopic expression was observed upon differentiation efficiency at 37°C or 20°C.

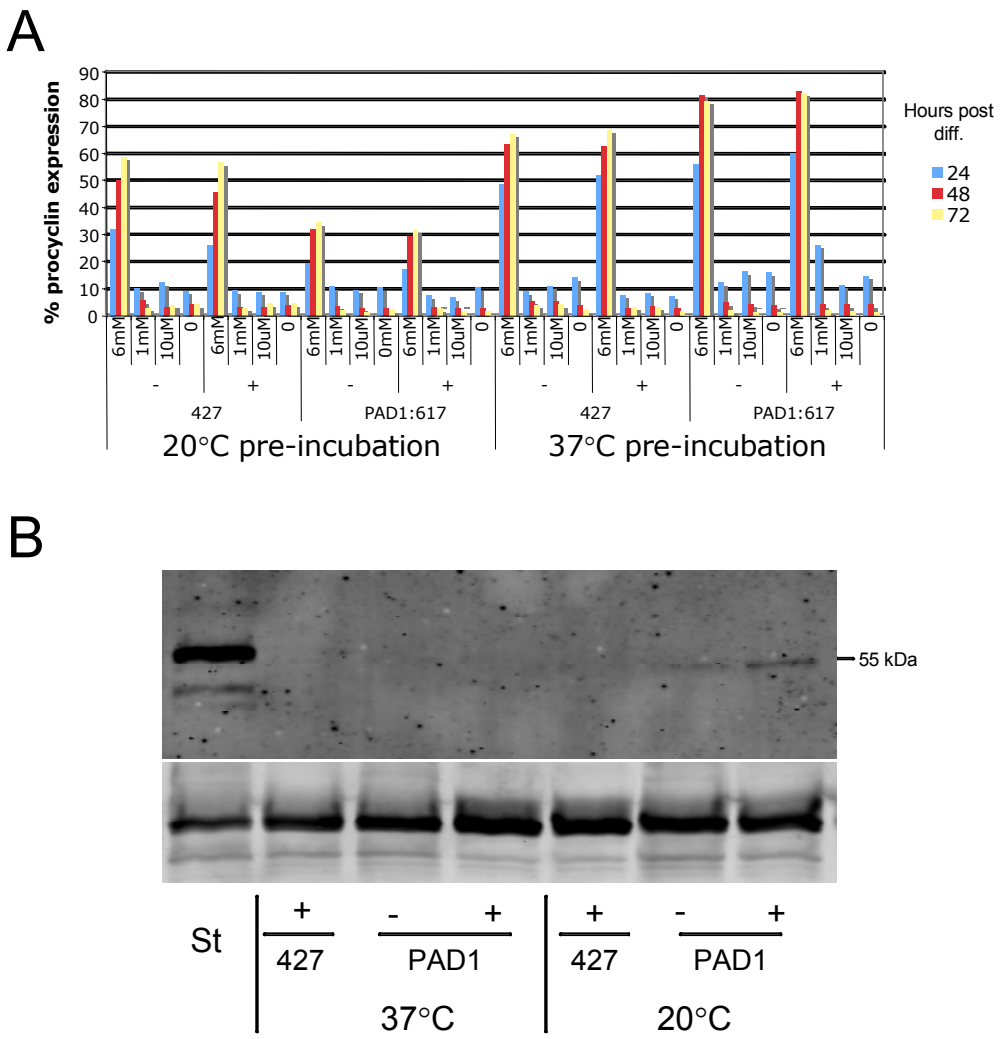


Figure 6.1-1 The effect of ectopic expression of PAD1 in combination with cold shock upon the sensitivity to cis-aconitate. (A) Monomorphic cells were assayed for their sensitivity to *cis*-aconitate using ectopic expression of PAD1 and cold shock as variables. Differentiation was assayed by flow cytometry analysis of EP expression over 72 hours. (B) PAD1 protein expression was compared with stumpy form cells by Western blot using the α-PAD1 antibody. PAD1 was not detected in monomorphic cells at 37°C in any of the conditions. When cold shocked, expression was detected in the PAD1:617 cell line, and shown to be 3 fold upregulated in induced cells, but nonetheless 10 fold lower than was detected in stumpy form cells. ± refers to tetracycline induction, St = stumpy form, 427 = non-transfected, PAD1 = 427 cells containing the PAD1:pHD617 construct.

6.1.2 The affect of incubating cells with different concentrations of *cis*-aconitate during cold shock and ectopic expression of PAD1

Hitherto, the detection of *cis*-aconitate has been limited to observations of differentiation *after* cells have been subjected to cold shock (as described by Engstler *et al.*(2004)). However, in this context it was possible that incubating the trypanosomes with different concentrations of *cis*-aconitate *during* cold shock would be more revealing. It may be that the cold shock effect is rapidly reversible and as such hypersensitivity would not be observed by subsequent incubation at 27°C.

To investigate this possibility, cells ectopically expressing PAD1 were incubated with different concentrations of *cis*-aconitate at 27°C and under cold shock conditions; non-induced cells and cells that did not have the ectopic expression cassette were included as experimental controls (see Figure 6.1-2A). The level of differentiation was measured after 24 hours by assaying for the expression of EP procyclin using flow cytometry (see Materials and Methods).

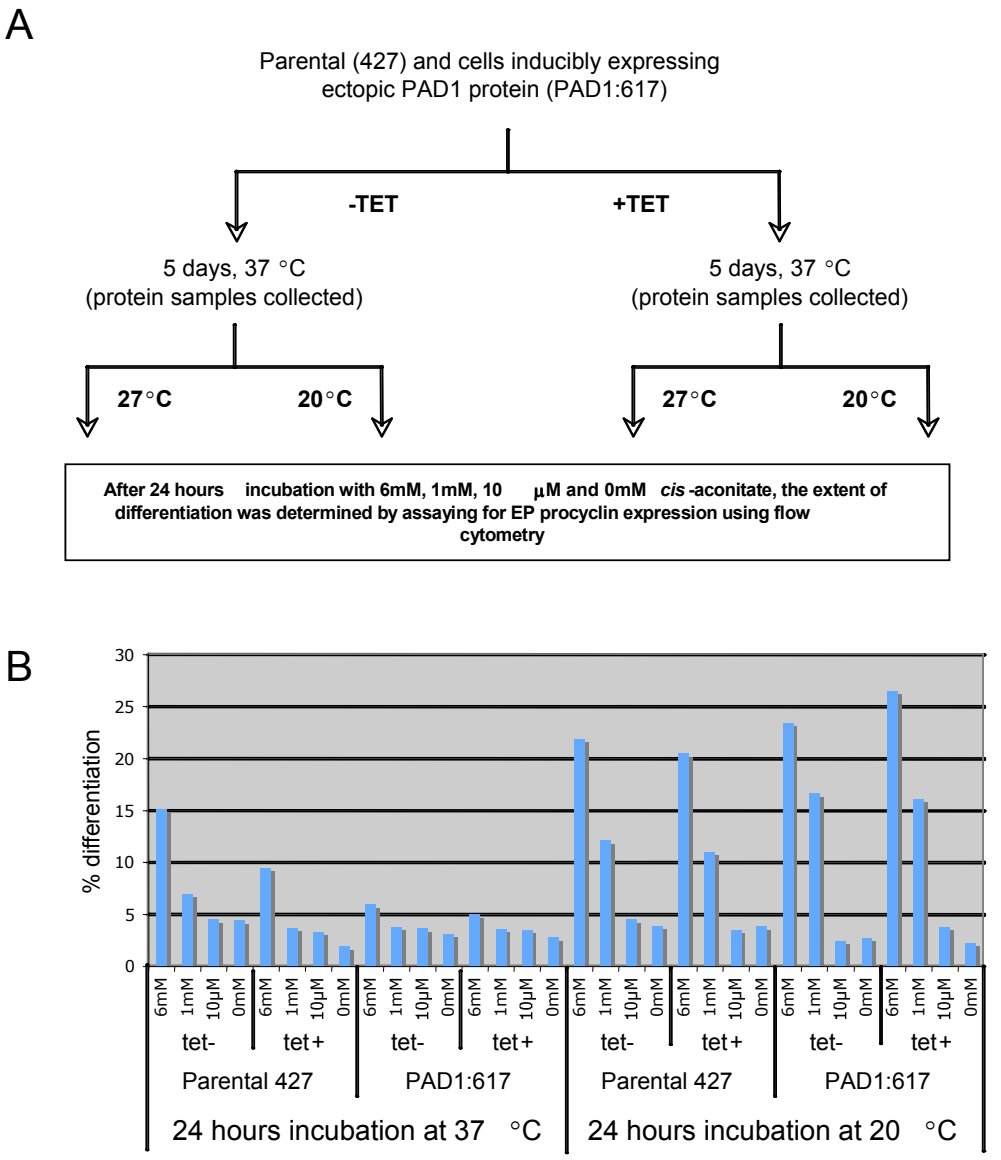


Figure 6.1-2 The effect of incubating *cis*-aconitate with cells ectopically expressing PAD1 during cold shock. (A) Parental cells and PAD1:pHD617 cells were incubated with, or without, tetracycline for 5 days prior to the addition, or not, of different concentrations of *cis*-aconitate at either 37°C or 20°C. (B) Differentiation was assayed by flow cytometry analysis of EP procyclin expression after approximately 24 hours. There was a mild, inducible, increase in differentiation capacity at 6 mM, and a non-inducible increase in differentiation capacity at 1 mM cis-aconitate. A reduced capacity for differentiation was observed at 37°C. Interestingly, cells responded better to *cis*-aconitate at 20°C irrespective of PAD1 induction.

Figure 6.1-2B shows that, in cells pre-incubated at 37°C, parental Lister 427 cells appear to differentiate better in response to 6 mM *cis*-aconitate than cells undergoing the ectopic expression of PAD1. However, this effect was not inducible and so may be unrelated to PAD1 expression. Similarly, no enhanced sensitivity to *cis*-aconitate was observed for the PAD expressing line and only background levels of EP expression were detected at 1 mM, 10 μM and in the absence of *cis*-aconitate.

Differentiation was more efficient at 20°C than 37°C for both of the cell lines. Surprisingly, both the parental cell line and the cell line containing the ectopic copy of PAD1 respond to *cis*-aconitate at 1 mM well above background levels - this effect was not observed when *cis*-aconitate was added after cold shock (see Figure 6.1-1). Additionally, the cell line containing the ectopic PAD1 has higher differentiation capacity than control cells at 6 mM and 1 mM; however, at 1 mM, where any effect of PAD1 expression upon the sensitivity to *cis*-aconitate should have been most noticeable, this effect was not inducible and was probably unrelated to PAD1 expression. Moreover, no increase in differentiation was observed at 10 μM, showing that these cells did not become hypersensitive to the low concentrations of *cis*-aconitate reported by Engstler *et al*. (2004).

Note that this experiment was performed alongside that described in 6.1.1, and that expression of PAD1 was not detected in the cell lines maintained at 37°C (see Figure 6.1-1B, 37°C, - and + PAD1), but was detected in those that were equivalently cold shocked.

Thus we can conclude that, although monomorphic cells may respond to *cis*-aconitate more efficiently under cold shock conditions, hypersensitivity to *cis*-aconitate was limited to stumpy form cells and that increased expression of PAD1 has not conferred this stumpy form characteristic. Nonetheless, in this experiment, PAD1 expression combined with cold shock may have stimulated a mild, enhanced sensitivity to *cis*-aconitate at 1 mM and 6 mM. It would be interesting to repeat this experiment with ectopic expression of PAD2; however, the upregulation of PAD2 in

response to cold shock was shown relatively late in this study, and time constraints prevented a repeat of this analysis.

6.2 Using RNA interference to ablate PAD1 in monomorphic slender cells

RNAi in *Trypanosoma brucei* was originally discovered by accident when trypanosomes were transfected with a construct capable of forming a hairpin loop with tubulin RNA (Ngo *et al.* (1998)). Since then, the technology has matured and inducible, heritable RNA interference of essential genes is possible (Shi *et al.* (2000)) (LaCount *et al.* (2000)) (Wang *et al.* (2000)) (DaRocha *et al.*(2004)). Two types of RNAi construct are in use. The first contain opposing promoters whereby a nucleotide sequence homologous to the target mRNA transcript is cloned between the two opposing promoters such that, when both promoters initiate transcription, double stranded RNA will be synthesised. The second type are stem loop vectors, whereby two duplicate homologous sequences are cloned either side of a non-specific spacer sequence downstream of a promoter such that they are inverted. When transcription is initiated at the promoter, a stem-loop structure is synthesised whereby the stem is double stranded RNA specific to the target mRNA transcript. Several variants of these constructs are in use, whereby the type of promoter (i.e. procyclin, ribosomal RNA, VSG and T7) and the site of genomic integration (commonly the ribosomal spacer or the minichromosome) have been modified. Additionally, the promoter driving transcription usually contains a tetracycline repressor protein (TetR) binding sequence (tetracycline operator) such that when the TetR is expressed, expression of double stranded RNA is only possible in the presence of tetracycline.

6.2.1 Generating the RNAi construct

The stem loop RNAi construct has been noted to cause more efficient knock down of target transcripts as well as having tighter regulation (Professor James Bangs, University of Wisconsin, personal communication). Due to the homologous nature of the PAD array, there were very tight restrictions on which regions of PAD1 could be used as a target. The theoretical minimal requirement for the RNAi process to

work is predicted to be 21 nucleotides (Zamore *et al.* (2000)) (Elbashir *et al*. (2001)); indeed, synthetic 21 nucleotide dsRNA corresponding to the size of small interfering RNA (siRNA) has been shown to ablate a target transcript (Elbashir *et al.* (2001)). However, the minimal size of target sequence homology within an RNAi stem loop vector for efficient gene ablation to occur has not been experimentally determined.

Nonetheless, a 76 base-pair nucleotide sequence was amplified from trypanosome genomic DNA by the polymerase chain reaction. The forward primer (primer #108) contained HindIII and BclI (5' – 3') restriction enzyme recognition sites at the 5' end, and the reverse primer (primer #109) contained XhoI and NdeI sites. The amplicon was purified and digested in two different restriction digest reactions: (1) XhoI and BclI, and (2) HindIII and XbaI, and re-purified. Number 1 insert was ligated with XhoI and BamHI digested pALC14 and used to transform XL1Blue cells; colonies were screened by restriction digest with BamHI and XhoI (plasmids containing insert number 1 did not release an insert because the BamHI recognition site was destroyed). A positive colony was cultured and a small scale preparation of DNA was made for subsequent HindIII and XbaI restriction digestion and ligation with digested insert number 2. After transformation of XL1Blue cells, resulting colonies were screened for the presence of both inserts in inverse orientation by the polymerase chain reaction with the forward primer only.

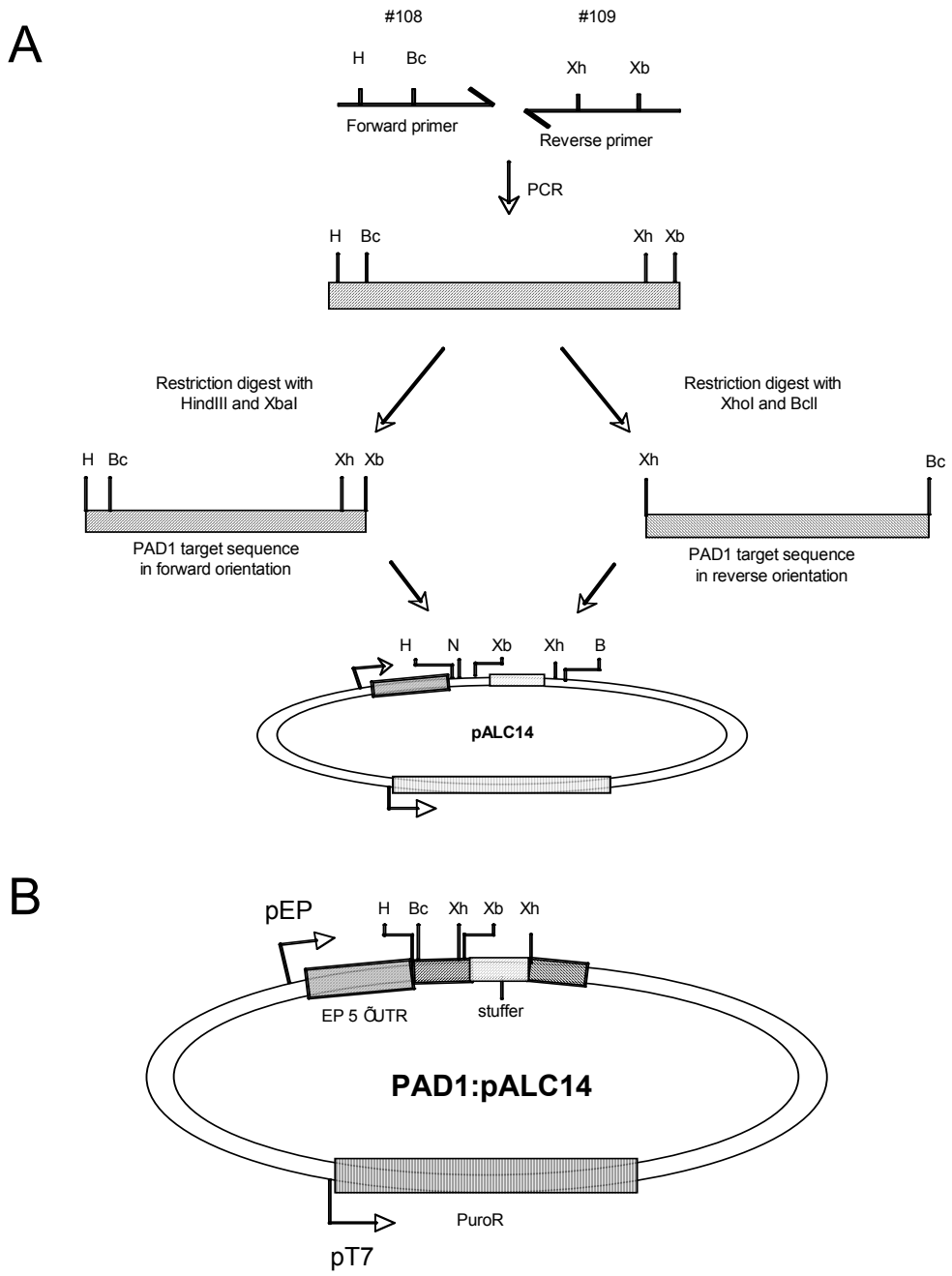


Figure 6.2-1 Making the PAD1:pALC14 RNAi DNA construct. (A) Primers #108 and 109 were used to generate a 76 nucleotide amplicon (5' → 3' directionality is indicated by the arrow). The amplicon was digested with either XhoI and BclI, or HindIII and XbaI, and cloned into either XhoI and BamHI, or HindIII and XbaI digested pALC14 respectively. (B) A schematic of the relevant regions of PAD1:pALC14 plasmid to knock down PAD1 is shown. Note that the plasmid BamHI and the amplicon BclI sites were destroyed upon cloning. Bc: BclI, B: BamHI, H: HindIII, Xh: XhoI, Xb: XbaI.

6.2.2 The effect of inducing RNAi upon PAD1

The monomorphic S16 SMB cell line (Wirtz *et al*. (1998)) expressing T7 polymerase and the tetracycline repressor was transfected with the PAD1 RNAi plasmid PAD1:pALC14. Cells were induced, or not, to knock down PAD1 mRNA with tetracycline for 4 days and subsequently exposed to *cis*-aconitate and a drop in temperature to 27°C. Induced and non-induced untransfected S16 cells were included as negative controls. RNA and protein samples were made prior to *cis*-aconitate treatment to determine the extent of knock-down of PAD1.

Figure 6.2-2A shows that there did not appear to be detectable RNA ablation in response to tetracycline in the PAD1:pALC14 cell line. Although the RNA was slightly degraded, there was a band corresponding to PAD1 in cells incubated with or without tetracycline. As expected, the PAD1 transcript was present in far higher levels in the stumpy form positive control. An overexposure is shown because the PAD1 transcript is present at very low levels in monomorphic cells.

It appears that RNAi on PAD1 was unsuccessful in this experiment. The reasons for this are unclear; it may be that knockdown of PAD1 was lethal such that generating a cell line containing PAD1:pALC14 selects for cells that are unable to execute RNAi. However, given that PAD1 mRNA is present in very low levels, and that PAD1 protein is undetectable in monomorphic slender cells, this is unlikely. It is more likely that 76 base-pairs of nucleotide sequence was insufficient to generate effective RNA ablation.

Although there was no clear evidence of PAD1 mRNA ablation, differentiation capacity was assayed at 0, 24, 48 and 72 hours post *cis*-aconitate treatment by FACS analysis of EP procyclin expression in PAD1 RNAi and parental cell lines with and without incubation with tetracycline (see Figure 6.2-2B). As expected, no differentiation phenotype was observed in response to tetracycline.

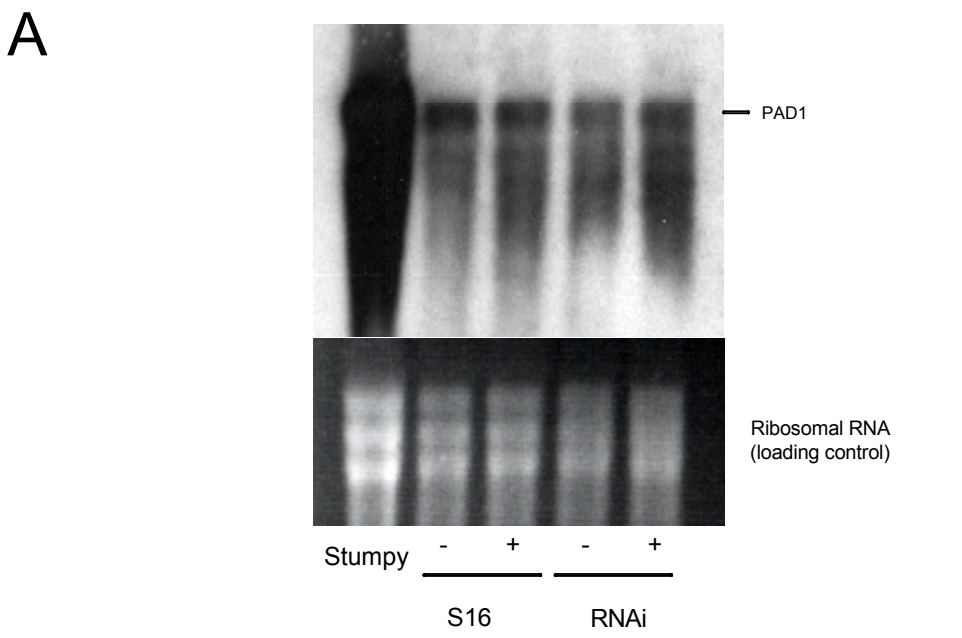


Figure 6.2-2 The effect of targeting PAD1 by RNAi for differentiation in monomorphic cells. (A) A Northern blot to determine if PAD1 mRNA ablation was successful. No ablation of the low level PAD1 transcript was detected. The blot was overexposed to allow visualization of the of the low level PAD1 mRNA in monomorphic S16, SMB cells. (B) The effect of PAD1 RNAi induction upon the capacity of monomorphic slender cells to respond to 6 mM *cis*-aconitate at 27°C. No effect upon differentiation was detected.

6.3 Using RNA interference to ablate the PAD array

Discounting technical issues, there are some potential biological caveats associated with phenotypic analysis of PAD1 mRNA ablation. PAD1 shares its RNA expression profile with PAD5 and PAD7 and there is a strong likelihood of functional overlap between the PAD proteins due to their high degree of homology. Thus, ablation of one member in the PAD array may be compensated for by the presence of other PAD proteins resulting in no observable phenotype. For this reason it was of interest to ablate every member of the PAD array at once. The high degree of homology allows all eight genes in the PAD array to be ablated when the entire coding sequence of PAD1 was expressed as double stranded DNA. Moreover, this was predicted to overcome possible issues of insufficient length of double stranded RNA that may have been associated with the PAD1 RNAi (see 6.2).

6.3.1 Generating the RNAi construct

The entire PAD1 coding region (approximately 1800 base-pairs) was amplified from trypanosome genomic DNA by the polymerase chain reaction using a forward primer (primer #106) containing HindIII and BclI (5' – 3') restriction enzyme recognition sites at the 5' end, and a reverse primer (primer #109) containing XhoI and NdeI sites. Cloning followed the same strategy as cloning PAD1:pALC14 (see Figure 6.2-1) except for the screening steps: successful insertion of the first insert was screened by restriction enzyme digestion with BamHI and HindIII, this releasing a 346 base-pair fragment due to the internal BamHI site, whereas BamHI digestion of the final construct released a 1.1 Kb fragment. The resulting plasmid was named "array:pALC14". All primer sequences are listed in Appendix A.

6.3.2 The affect of inducing RNAi upon PAD protein and RNA

Monomorphic slender S16 SMB cells were transfected with array:pALC14. The resulting cell line was cultured in HMI9 media in the presence, or absence, of tetracycline for 5 days prior to preparation of protein and RNA (see Materials and

Methods for details). The parental S16 SMB cells with and without tetracycline were included as controls.

A Northern analysis was performed on the RNA using a DIG labelled riboprobe designed to detect every member of the PAD array. The RNA appeared to have degraded somewhat, probably due to RNase mediated deterioration, but there was a clear reduction of signal detection in the cell line transfected with array:pALC14 when incubated with tetracycline (see Figure 6.3-1A) in comparison with those with no tetracycline, and the S16 SMB parental cell line. This indicated that the RNAi was likely successful.

A Western analysis was not thought to be necessary as previous Western blots using both the α-PAD1, α-PAD2 and the α-PAD-array antibodies did not detect any PAD family protein expression in monomorphic slender cells (see Sections 5.4, 5.5 and 5.6).

6.3.3 The affect of inducing RNAi upon capacity for differentiation

Subsequent to tetracycline induction, the four cultures were induced to differentiate by supplementing HMI9 media with 6 mM *cis*-aconitate and a reduction in temperature from 37°C to 27°C. Flow cytometry samples prepared at 0, 24, 48 and 72 hours post *cis*-aconitate addition for analysis of differentiation by EP procyclin expression (see Materials and Methods for details).

Figure 6.3-1B shows that no difference was detected in differentiation capacity between the different cell lines, irrespective of the presence of tetracycline. Thus it appears that ablation of PAD transcript did not affect the capacity of cells to differentiate. If any of the PAD genes have a role in differentiation in monomorphic slender cells it would appear that they are able to exert their effect at very low levels of protein. However, given the lack of detectable PAD family protein expression in monomorphic slender cells, and the lack a differentiation phenotype in response to RNAi mediated ablation of the PAD family transcripts, it appears likely that monomorphic slender cells do not respond to *cis*-aconitate through a PAD protein mediated mechanism.

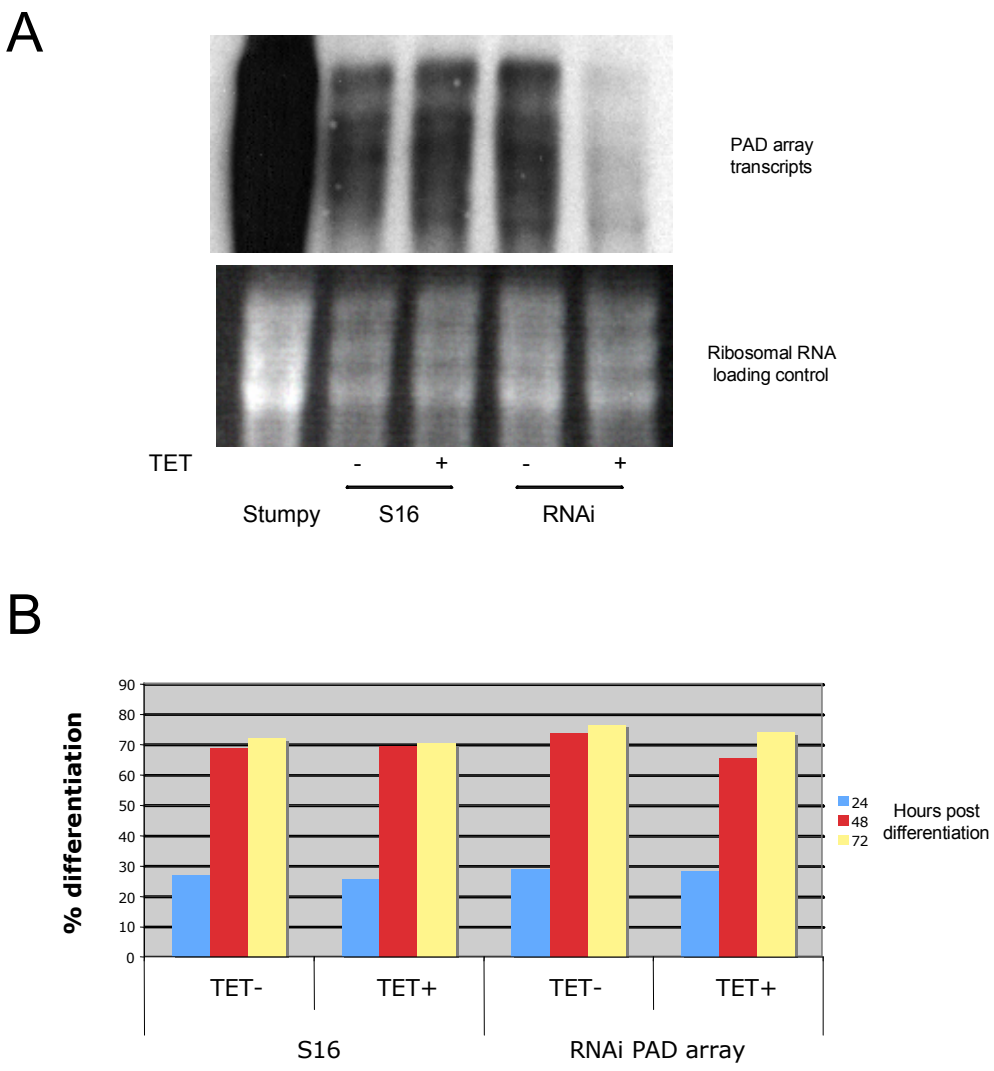


Figure 6.3-1 The effect of ablating the mRNA transcripts of the entire PAD array. Cell lines were generated that would ablate all of the PAD mRNA transcripts upon induction with tetracycline and the affect upon differentiation was assayed. (A) A Northern analysis to determine if PAD1 mRNA ablation was successful. Note that the mRNA is partially degraded, and that the blot was overexposed in order to detect the low level PAD family transcripts clearly. (B) The differentiation capacity of RNAi and parental cell lines was assayed in the presence and in the absence of tetracycline over a 72 hour timecourse. There did not appear to be any differentiation phenotype in response to RNAi induction, and the RNAi cell line differentiated with similar kinetics to the untransfected parental cell line.

6.4 Generating transgenic knockout cell lines without the PAD1 gene

RNAi ablation of PAD1 was unsuccessful, and RNAi ablation of the array transcripts was not complete. Therefore, a gene knockout strategy was undertaken.

The most desirable cell line in which to perform this knock out is a pleomorphic cell line capable of forming all stages of the trypanosome lifecycle. PAD1 is highly expressed in the stumpy form, thus it is logical to determine the affect upon differentiation of removing PAD1 in stumpy form cells. However, pleomorphic cells are notoriously difficult to culture and although initial attempts were successful, sustained culturing failed. For these reasons the knockout was attempted in monomorphic cells. This is far from ideal; monomorphic cells are a laboratory artefact that does not differentiate synchronously, as is the case with stumpy form cells. Moreover, as shown earlier, PAD protein was undetectable when monomorphic cells were analysed using either the α-PAD1 antibodies or the α-array antibodies, and no upregulation was detected in response to *cis*-aconitate or at any point through differentiation. This raises the possibility that either (A) PAD proteins do not play an essential role in the differentiation of monomorphic cells or (B) the PAD proteins exert their effect at levels undetectable by Western blot. If deletion mutants exhibited a differentiation phenotype, this would form a strong argument in favour of PAD proteins transmitting the signal to differentiate in monomorphic cells.

6.4.1 Making the constructs to remove PAD1

Generating knockout cell lines has some technical differences in the design of the DNA construct when compared to other types of genetic manipulation. DNA constructs designed to integrate into the genome use regions homologous to the target site (for example the ribosomal spacer). Upon transfection into *T. brucei*, the cellular machinery recognises the homologous, double stranded ends, and aligns the ends of the fragment with their target sequence to 'repair' the break and integrate the construct; the region of homology is therefore duplicated and the DNA construct is inserted in the reverse orientation. A gene knockout, however, is designed such that

a linear fragment of DNA is generated that contains a resistance cassette flanked by terminal regions of homology to sequences flanking to the gene to be removed. The DNA repair mechanism of the cell aligns the terminal regions of homology, and replaces the target gene with resistance cassette. Transgenic cells can be selected for through drug resistance, resulting in a final transfection efficiency of 1 transgenic cell per 10^5 to 10^7 wild type cells. *T. brucei* is a diploid organism meaning that both alleles of a gene must be removed and, conventionally, a conditional rescue copy of the gene is inserted under the control of an inducible promoter to verify that any phenotype observed is due to the removal of the target gene. Hence, the DNA manipulations involved are quite complex.

The DNA constructs to remove the PAD1 gene were designed using 2 regions of homology upstream of the PAD1 gene, and 2 regions downstream such that 2 of the sequences were nested within the other two (see Figure 6.4-1A). Each region of homology was designed to be approximately 250 – 300 nucleotides long and was designated A, B, F and E (starting from the most upstream region). Region A and E were amplified and cloned into the cloning sites flanking the neomycin and TetR cassettes (referred to as AE-neo), and B and F were cloned into restriction sites flanking a hygromycin resistance cassette (referred to as BF-hygro). The resulting DNA constructs were verified by PCR and sequencing to confirm that the regions of homology were present in the correct orientation (the cloning primers, a schematic of the deletion constructs, and the genomic regions targeted by the deletion constructs, are shown in Appendix A).

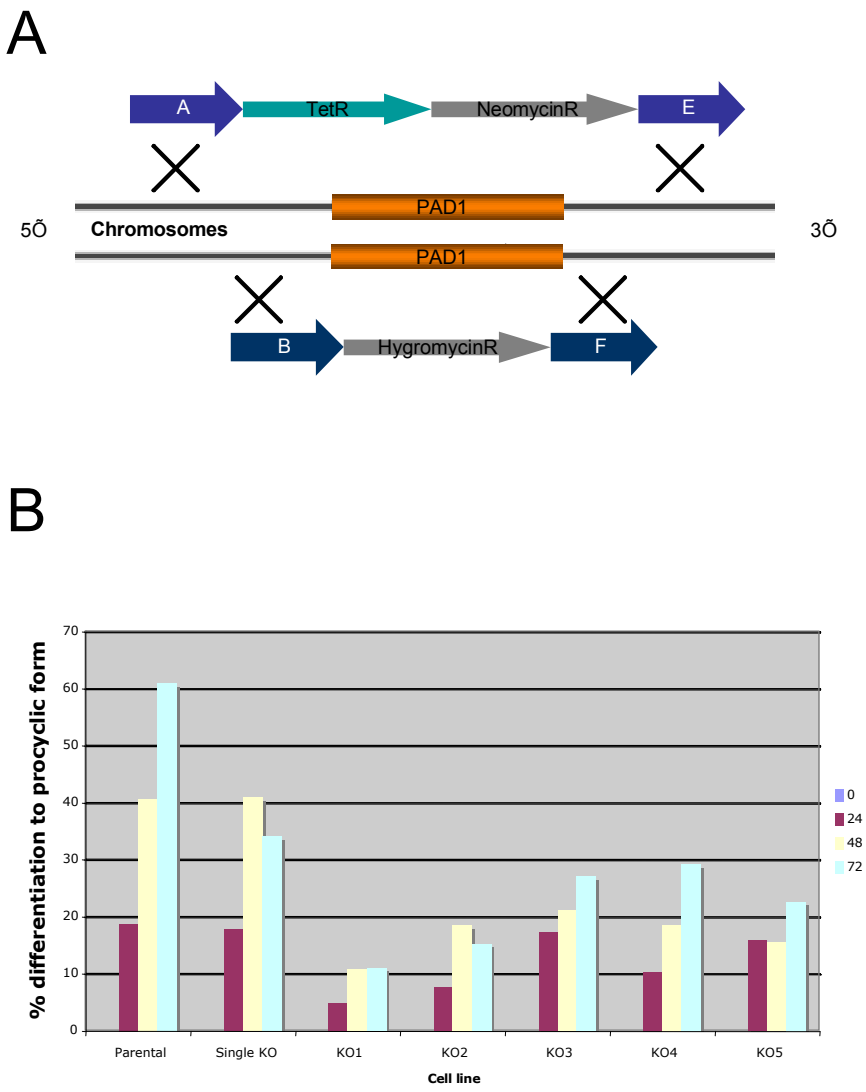


Figure 6.4-1 Design of the PAD1 knockout constructs and the differentiation capacity of a cell line containing them. (A) DNA constructs were designed using regions of homology flanking the PAD1 gene. (B) Cell lines that had been transfected with these constructs were assayed for their differentiation capacity. 5 semi-independent cell lines all show impaired differentiation analysed by immunofluorescence analysis of cell expressing EP procyclin. This analysis was repeated using cell line 1 and assaying differentiation capacity by flow cytometry, and showed a similar result.

6.4.2 Analysing the transgenic cell lines

T. brucei strain Lister 427 was transfected with AE-neo and the resulting cell lines were pooled prior to being transfected with BF-hygro. The resulting 5 semi-independent cell lines were subsequently assayed for a differentiation phenotype by culturing at 27°C in the presence of 6 mM *cis*-aconitate. Differentiating cells were fixed in methanol at 0, 24, 48 and 72 hours post-differentiation, and stained with EP procyclin to assay the percentage of cells that had differentiated by UV immunofluorescence microscopy (see Materials and Methods for details). The single knockout transgenic cell line showed no differentiation phenotype until 72 hours when compared with the parental Lister 427 cell lines; however, all 5 of the semi-independent double transgenic cell lines showed a strong differentiation phenotype in comparison with the parental Lister 427 strain. 'KO1' attains a maximum differentiation of approximately 12%, compared with more than 60% of the parental cells. Subsequent differentiation experiments using flow cytometry to assay EP procyclin expression re-capitulated this result (data not shown).

These data suggest that PAD1 does indeed have a role in differentiation of monomorphic slender cells, despite undetectable levels of PAD1 protein. This might lead one to speculate that PAD1 functions under these conditions at very low protein concentrations, thus explaining the lack of any differentiation phenotype when RNAi mediated ablation is induced. The differentiation that was observed must therefore have been due to entry of *cis*-aconitate into the cell by other means.

In order to introduce a rescue copy of PAD1 into the hygromycin and G418 resistant KO1 cell line, pHD617(hygro) was modified to replace the hygromycin cassette with a puromycin resistance cassette. Briefly, the 5' end of the hygromycin open reading frame, along with the upstream rRNA promoter and the actin 5' UTR was removed by restriction enzyme digest with XhoI and NdeI. An amplicon containing the rRNA promoter with a 5' XhoI site and a 3' NdeI site was then inserted into the agarose purified vector backbone to facilitate subsequent insertion of an amplicon containing the puromycin cassette; the puromycin cassette was PCR amplified from pHD1034

(Quijada *et al.* (2002)) and inserted into the NdeI and NheI restriction sites of the backbone, thus creating pHD617(puro). This analysis was performed prior to the availability of the α-PAD1 antibody, and therefore a Ty epitope tag was incorporated into the C terminus of PAD1, to make pHD617(puro):PAD1Ty.

Figure 6.4-2A shows that the introduction of an ectopically inducible rescue copy of PAD1-Ty into cell line KO1 was able to partially rescue the differentiation phenotype, but, surprisingly, did not restore full parental differentiation capacity and was not inducible. The lack of inducibility may have been due to leaky expression allowing production of low, but functional, levels of PAD1. However, the lack of complete rescue was more difficult to explain, but one explanation was that the presence of the Ty epitope tag was interfering with PAD1 function.

A Western blot was performed using the BB2 antibody to determine whether expression of the epitope tagged PAD1 was leaky, thus explaining the partial rescue of the phenotype in the absence of tetracycline. Figure 6.4-2B shows that 2 strong bands were detected in the rescue cells when they were incubated with tetracycline, and 1 much weaker band was detected when no tetracycline was added. However, this weaker band may have been a cross reactive band, as this was also detected in the knockout cells that had not been transfected with the epitope tagged gene. This blot was difficult to interpret fully as some of the signal was obscured.

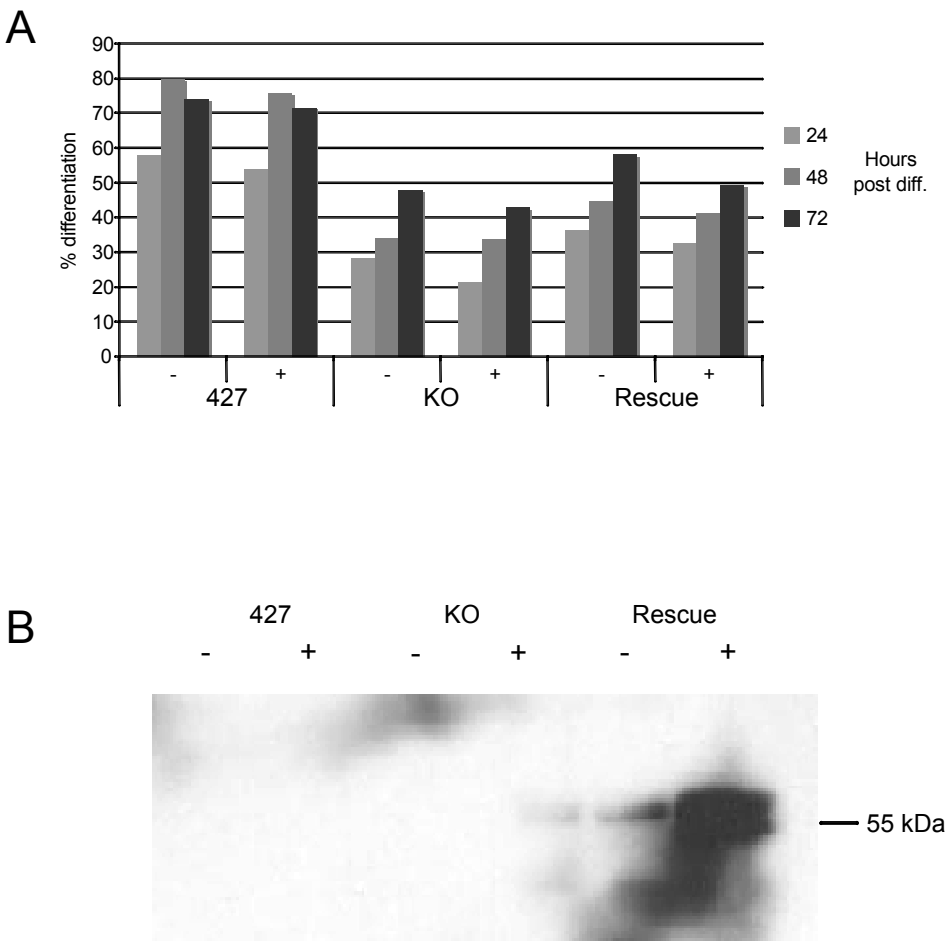


Figure 6.4-2 Rescue of PAD1 knockout cell lines. An epitope tagged rescue copy of PAD1 was integrated into the PAD1 knockout cell line in order to rescue the phenotype. (A) Rescue of the phenotype was successful, but incomplete and not inducible. (B) A Western blot using the BB2 anti-Ty antibody shows that there did appear to be an inducible band. However, several unidentified bands were also detected.

A cleaner Western blot than Figure 6.4-2B would have been desirable. However, in the process of verifying the PAD1 null mutants, a PCR reaction using genomic DNA from the KO1 cell line as template and primers specific to PAD1 showed a strong amplicon at the size expected if PAD1 was still present within the genome. Further experiments showed this band in the other null mutant cell lines (see Figure 6.4-3). These bands were cloned into pGEM T easy and sequenced to verify that the PCR reaction has indeed amplified PAD1 and not another member of the PAD array; sequence data confirmed the presence of PAD1. Subsequent PCR reactions verified the presence of the hygromycin and neomycin cassettes, demonstrating that these cassettes have indeed integrated into the genome (see Figure 6.4-3).

Thus it appears that these cell lines retained the PAD1 gene and were therefore, in fact, not PAD1 deletion mutants. The reasons for this are unclear; it may be that the repetitive nature of the PAD array intergenic regions has resulted in a mal-integration of the DNA constructs, possibly into another area of the array.

Alignments of the targeting sequences with the PAD array intergenic regions (see Appendix A) showed that target sequence (TS) B contains 67 nucleotides of sequence from intergenic element I at its 3' end, which is present in the 5' UTR of all of the PAD genes. Nonetheless, it was not thought that this would interfere with the removal of PAD1 because there are 148 nucleotides upstream of this that are not present elsewhere within the array; moreover, the 3' TS in this construct, F, is entirely specific to PAD1 and does not contain sequence that is repeated elsewhere in the PAD array. Thus, the target construct BF-hygro was predicted to remove PAD1.

The 5' region of homology ('A') in AE-neo is entirely specific to upstream of PAD1 and does not contain sequences homologous to other region of the array. However, the TS 'E' in this construct contains 25 nucleotides of sequences of intergenic element VI, which is present in the PAD5/7 3' UTRs, thus giving rise to the possibility that 'E' may target to these regions. However, in the event of this occurrence, we would expect that all of the DNA sequence between PAD1 – 5 or

PAD1 – 7 to be removed; this was clearly not the case as PAD1 was shown to still be present in these cell lines. Thus, the construct AE-neo was predicted to work.

A previous analysis of a 28 kb region of genome (Kramer *et al.*(2006)) showed that there were major differences between homologous regions of genome in different *T. brucei* isolates, especially when these regions contained repetitive elements. The authors show that that a 14.7 kb fragment of DNA containing 4 genes was duplicated in an isolate of MITat.1.2 that had undergone serial passage in culture, and speculate that the duplication event occurred during unequal crossing over between repetitive DNA sequences on homologous chromosomes. Another study concluded the variation in DNA content in chromosome I between the genome reference 927/4 strain, STIB247 and Lister 427 was concentrated in the sub-telomeres and tandemly repeated genes (Callejas *et al*. (2006)). Furthermore, the authors report that all tandemly repeated gene arrays on chromosome I showed variation in the number of duplicated genes between the strains used, and speculate that the amplification and reduction in size of these arrays is an ongoing process. Thus, the evidence shown in this paper would lead one to predict that the 427 strain of *T. brucei* does not contain the same number of PAD genes as the genome reference strain. Furthermore, the genome reference database annotation states that the 'Number of repeated genes in array uncertain, region may contain misassembly'. This may explain the failure to remove PAD1 from the genome.

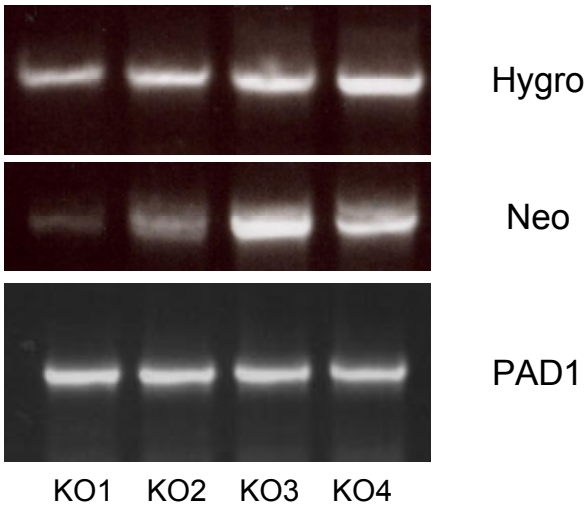


Figure 6.4-3 A PCR showing that PAD1, hygromycin resistance and neomycin resistance genes were still present in the PAD1 'null' mutants. A PCR reaction was preformed on PAD1 'null mutant' genomic DNA. Bands were obtained using primers specific to the hygromycin and neomycin resistance genes, as well as to PAD1. The bands from the PAD1 PCR reaction were subsequently cloned into pGEM T easy for sequencing to confirm they were indeed PAD1, and not a related PAD gene.

6.5 Generating transgenic PAD array null mutants

The homologous nature of the PAD genes makes it likely that there may be some overlap in function between the PAD genes. Thus, removing PAD1 may not show a phenotype due to rescue by other PAD members. For this reason it was decided to attempt to remove all 8 genes from the genome. This involved the replacement of approximately 27 kb of genomic DNA with a resistance cassette, making this the largest attempted knockout in a trypanosomatid. It may be that removing a segment of this size is not feasible by homologous recombination, so to maximise the chances of success, transfection efficiency was increased using the Amaxa nucleofector machine and buffers. The AMAXA protocol was reported to increase stable transfection efficiency by a factor of approximately 1000 fold (Burkard *et al.* (2007)) and has been used frequently with high success rates.

6.5.1 Generation of PAD array deletion constructs

The previously described constructs AE-neo and BF-hygro (see Section 6.4.1) were modified to replace the PAD1 3' target sequences E and F with regions downstream of PAD8 (designated 'C' and 'D', where D is upstream of C), such that two constructs were made that were made designed to remove the entire array ('AC-neo' and 'BD-hygro'). The BF-hygro target sequences were nested within the AC-neo target sequences, such that transfection of a cell line where 1 copy of the array has been removed by AC-neo with BF-hygro will remove the other copy of the array. Attempts to obtain transfectants with these constructs were unsuccessful, however, possibly due to the size of the region of genomic DNA to be replaced.

It was decided to make DNA constructs with larger regions of homology in order to increase the chances of success. A region of approximately 2.8 kb upstream of PAD1 (designated 'G') was cloned in place of A, and a region 1.6 kb downstream of PAD8 (designated 'H') was cloned in place of E in the AE-neo DNA construct, thus generating a construct (designated GH-neo) with very large regions of homology to the regions flanking the PAD array. Due to the large nature of these regions of

homology, a nested knockout was not possible. Therefore, the construct designed for the second round transfection was identical to GH-neo except that the neomycin resistance cassette was replaced with a hygromycin resistance cassette, thus generating 'GH-hygro'. Schematics of the deletion constructs are shown in Appendix A.

Lister 427 cells were transfected with GH-neo using the Amaxa nucleofection protocol (see Materials and Methods). Resulting cell lines were pooled and re-transfected with the GH-hygro cassette.

6.5.2 Analysing the transgenic cell lines

Given the failiure to generate PAD1 deletion mutants, it was a priority to verify the presence of the PAD genes in cells transfected with GH-neo and GH-hygro. Genomic DNA from the cell line containing GH-neo and GH-hygro was used as template in a PCR reaction to determine the presence of PAD genes. The primers used in the reaction were designed to bind to any of the PAD genes and generate a 400 nucleotide amplicon if PAD genes were still present within the genome. A band corresponding to the expected size of a PAD specific amplicon was generated by this PCR reaction; cloning and sequencing determined that this band was in fact PAD specific, thus the cell lines still retain PAD genes and are therefore not PAD array knockout cell lines (data not shown). Differentiation assays were not performed as the data would have been impossible to interpret.

It was suspected that there was a common reason for the failure to remove both PAD1, and the PAD array from the genome. It may be that 1 or more of the PAD genes is essential in monomorphic slender cells; however, this would appear unlikely as PAD-array RNAi cell lines did not show a growth phenotype (data not shown), and PAD protein was not detected in these cells. As discussed (see Section 6.4.2), it appears likely that this was associated with possible differences between the genome reference strain and the 427 cell line used for this study, or a genome mis-assembly in the genome reference data.

It was of interest to perform further analyses on the 'pseudo' PAD1 and PAD array deletion mutants. For example, determining the integration sites of these constructs may help to determine whether they have integrated into the predicted position, and thus allow us to infer the presence of multiple copies of PAD1, and the PAD array. However, it was decided that it was of higher priority to implement an alternative strategy that would allow the role of the PAD genes in differentiation to be determined. This involved implementing an RNAi strategy in pleomorphic cells.

6.6 RNAi mediated ablation of PAD family transcripts in pleomorphic cells

It was considered of primary importance to continue with further functional studies in a naturally occurring, pleomorphic cell line using an RNAi based approach. This had the dual advantage of (A), being independent of genome data mal-assembly and PAD gene copy number and (B), analysing the effect of PAD gene ablation in a cell type where PAD proteins were detectable.

6.6.1 RNAi mediated ablated transfection of PAD family transcripts using transient transfection

The vast majority of functional studies in *Trypanosoma brucei* are performed in monomorphic slender cell lines due to their ease of culture and genetic tractability. Pleomorphic cell lines can be very difficult to culture adapt and maintain in culture, possibly due to their tendency to enter cell cycle arrest at high culture densities. For this reason, initial PAD family functional analyses in stumpy form cells were performed by transient transfection of pleomorphic trypanosomes with a PAD array RNAi construct (see 6.3.1), such that expression of a stem loop structure was driven by an EP procyclin promoter and the PAD family mRNA transcripts were targeted for degradation.

Figure 6.6-1 shows a diagram depicting the design of the experiment. A mouse was inoculated with a pleomorphic parasites, and the cells harvested from the infection by DEAE cellulose anion exchange after several days (see Materials and Methods for details). Once harvested, the cells were transfected with either the stem loop PAD

array RNAi construct, or with water. After 24 hours recovery at 37°C to allow expression of the stem loop vector and slender to stumpy form differentiation, the cells were either maintained at 37°C, or incubated at 20°C overnight. Cells were subsequently incubated at 27°C in different concentration of *cis*-aconitate to assay their sensitivity to *cis*-aconitate and the effect of PAD array RNAi (see Materials and Methods for details).

A number of variables were considered in the design of this experiment. In particular, it was necessary to decide at what point in the slender to stumpy form transition to harvest cells from a parasitaemia and transfect with the RNAi constructs. If the cells were harvested too early during the transition, there may be insufficient time for them to subsequently differentiate to the stumpy form before the addition of *cis*-aconitate. However, if stumpy form parasites were transiently transfected, then PAD protein may have already been made prior to the initiation of RNAi thus preventing useful phenotypic analysis.

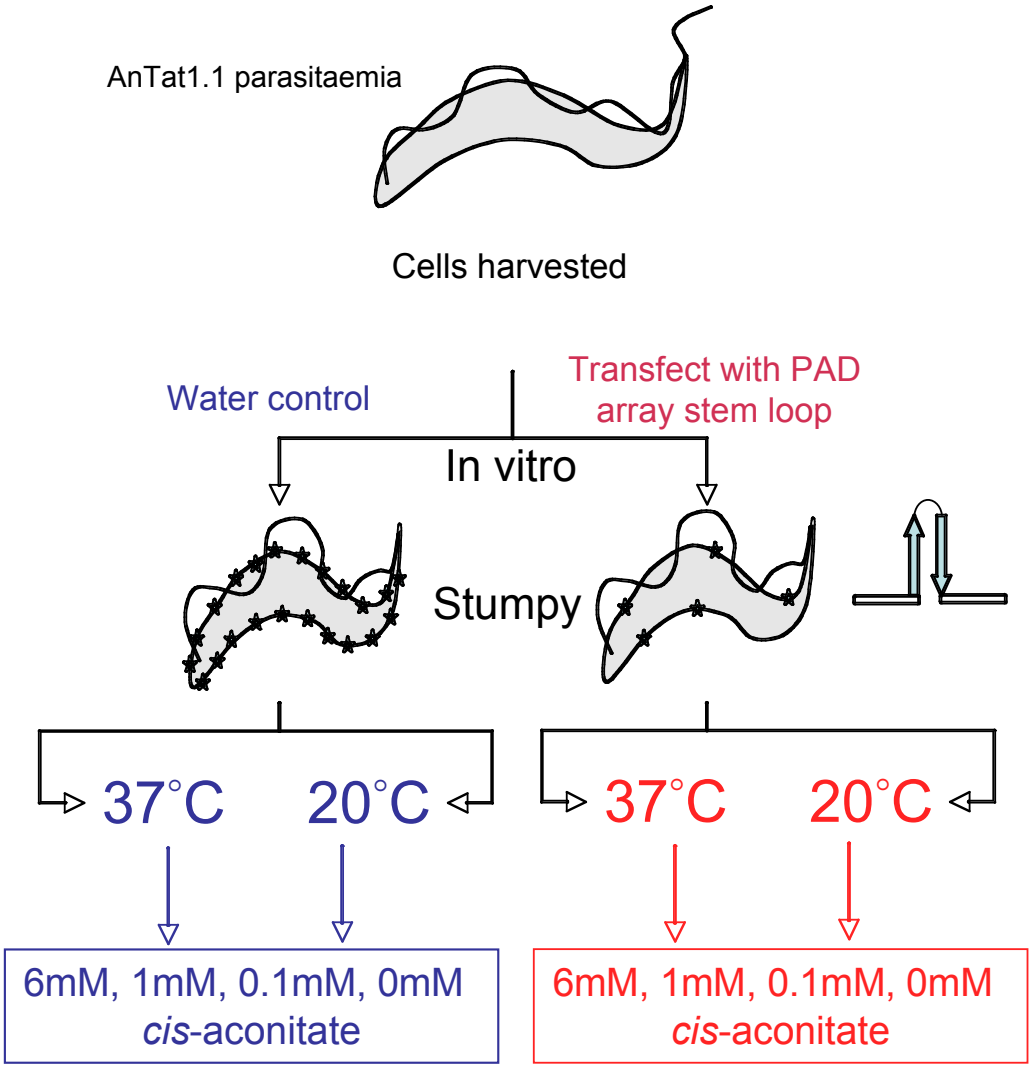


Figure 6.6-1 Transient RNAi in pleomorphic trypanosomes. Parasites from a mouse parasitaemia were harvested and transfected with the PAD array RNAi construct, or a water control. After 24 hours recovery, cells were either maintained at 37°C, or cold shocked at 20°C overnight. Cells were subsequently incubated at 27°C in different concentrations of *cis*-aconitate, and samples were prepared for flow cytometry over a timecourse to assay differentiation by expression of EP procyclin.

6.6.1.1 Transient transfection of trypanosomes from a mid-stage parasitaemia

An AnTat 1.1 mouse parasitaemia was closely monitored for the presence of slender, intermediate and stumpy form trypanosomes by taking blood smears from tail snips and observing the morphology of the parasites by light microscopy. At 5 days post-inoculation, an equal proportion of slender and intermediates was observed, with a small proportion (>10%) of stumpy forms, and the cells were transfected and assayed for their sensitivity to *cis*-aconitate at 37°C and 20°C as described (see Section 6.6.1 and Figure 6.6-1).

Figure 6.6-2A is a Western blot performed on protein prepared prior to the addition of *cis*-aconitate using the α-array antibody. There was a clear reduction in the levels of PAD protein in those cells that have been transfected with the RNAi construct; however, there remained a significant amount of PAD protein left in the cells.

Figure 6.6-2B showed that there was no effect on differentiation after RNAi when cells were incubated at 37°C before *cis*-aconitate addition. In contrast, when cells were incubated at 20°C before *cis*-aconitate addition, a mild reduction in differentiation was observed after RNAi at 1 mM *cis*-aconitate at 6 hours, although this was not observed at higher concentrations of *cis*-aconitate, or at later timepoints (Figure 6.6-2C). This may be due to experimental variation unrelated to PAD array RNAi, but it is worth noting that the 6 hour timepoint is most reflective of the level of initiation of differentiation, as the 24 hour timepoint is complicated by cell outgrowth due to re-entry of the cell into the cell cycle.

Interestingly, RNAi appeared to inhibit the cold shock induction of EP procyclin expression in these cells. This effect was observed at 0 hours where approximately 9% of RNAi cells expressed EP procyclin compared with 16% of non-RNAi cells, a >40% reduction.

Although further experiments may add statistical support to this observation, it was decided to repeat the experiment using cells from an earlier parasitaemia in order to

more completely reduce the PAD protein from the cells (see Section 6.6.1.2).

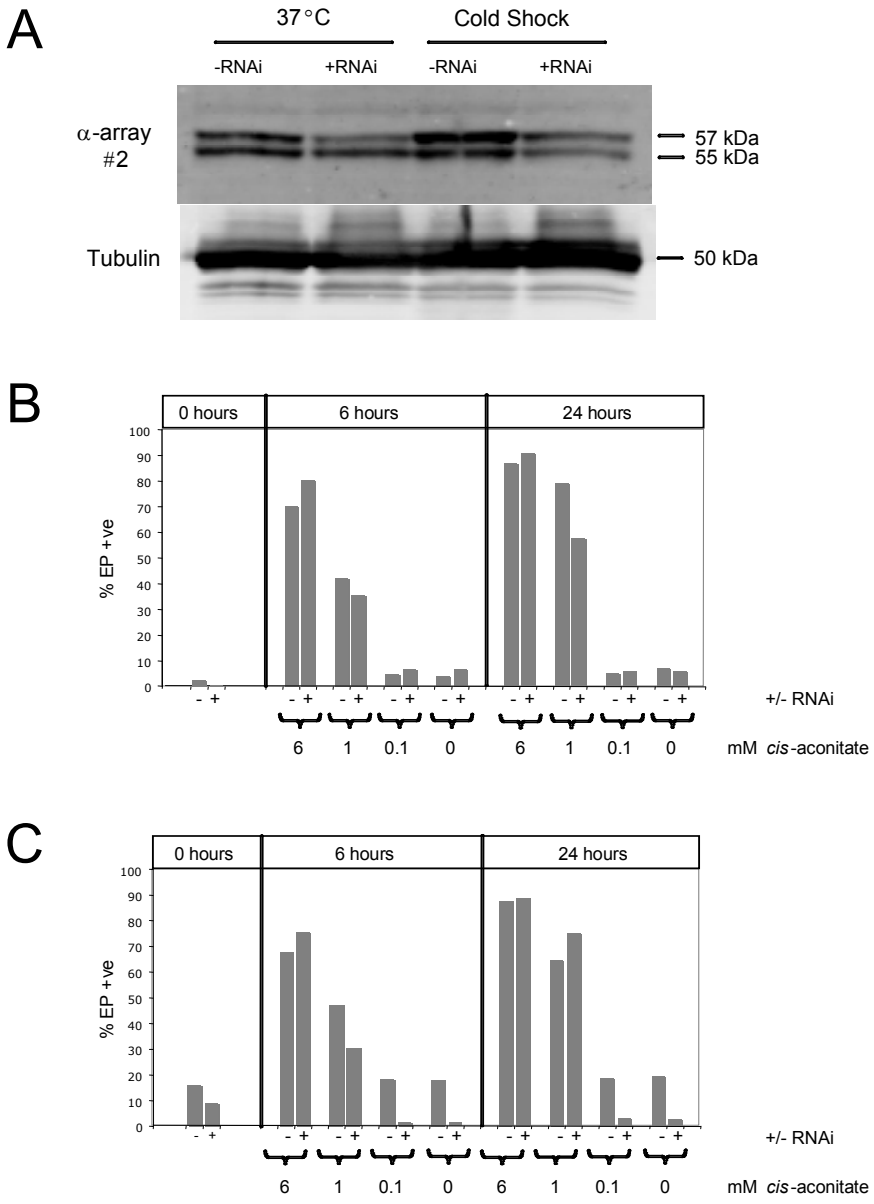


Figure 6.6-2 The effect of PAD family RNAi upon differentiation in morphologically pleomorphic cells. A pleomorphic population of cells were transiently transfected to ablate PAD family transcripts. (A) A Western blot shows that PAD RNAi resulted in the partial knockdown of PAD protein from transfected cells. (B) and (C) show the sensitivity to *cis*-aconitate when cells were pre-incubated at 37°C and 20°C respectively, as measured flow cytometry detection of EP procyclin expression. No convincing effect was observed at 37°C, but there may be an afect of RNAi at 1mM *cis*-aconitate after 6 hours under cold shock conditions. Interestingly, RNAi reduces induction of EP procyclin in response to cold shock.

6.6.1.2 Transient transfection of slender forms

As before (see Section 6.6.1.1) an AnTat 1.1 mouse infection was closely monitored for the presence of slender, intermediate and stumpy form trypanosomes by taking blood smears from tail snips and observing the morphology of the parasites by light microscopy. At approximately 3 days post-inoculation, the parasites were morphologically >99% slender, and the cells were harvested for transfection. RNAi treated and control cells were subsequently assayed for their sensitivity to *cis*-aconitate at 37°C and 20°C as described previously (see Section 6.6.1 and Figure 6.6-1).

Figure 6.6-3A shows a Western blot of RNAi treated and mock transfected cells that were incubated at either 37°C or 20°C prior to the addition of *cis*-aconitate. Consistent with stumpy form cell lysates, a 55 kDa and a 57 kDa band was detected using the α-array antibody in the mock-transfected cells, showing that these cells had differentiated to the stumpy form after treatment. In the cells subjected to RNAi, no PAD protein was detected despite cells becoming morphologically stumpy whilst in culture (data not shown), showing that RNAi was successful and highly efficient.

Figure 6.6-3B shows the analysis of 3 separate experiments. The average differentiation efficiency was assayed by flow cytometry at 0, 6 and 24 hours post *cis*-aconitate addition when cells were not previously cold shocked. Cells that were transfected with the RNAi construct showed an inhibition in their ability to respond to *cis*-aconitate when compared with equivalent negative controls. For example, at 6 hours post differentiation, 37°C, and 6 mM *cis*-aconitate, RNAi cells differentiate approximately 33% less efficiently than the negative controls. This was more pronounced at 24 hours (42%), possibly due to blood forms re-entering the cells cycle in the RNAi cells, or procyclic form cell death in the RNAi treated cells. This trend was repeated in cells incubated with 1 mM *cis*-aconitate, whereby cells undergoing RNAi differentiated 25% and 43% less efficiently at 6 hours and 24 hours post differentiation respectively when compared with mock-transfected cells.

Figure 6.6-3C shows means for the same experiments when cells were cold shocked prior to the addition of *cis*-aconitate. In this case, at 6 mM *cis*-aconitate there was a 31% and 47% reduction in differentiation at 6 hours and 24 hours, respectively. At 1 mM *cis*-aconitate this was more pronounced, with a 51% and 55% reduction at 6 hour and 24 hours post differentiation respectively.

A paired T test shows that cold shocked RNAi treated cells are impaired in their ability to respond to *cis*-aconitate with respect to the mock transfected controls, both at 6 hours and at 12 hours ($P < 0.005$). Non cold shocked, RNAi treated cells do not show statistically significant impairment at 6 hours ($P = 0.07$), but do at 12 hours ($P > 0.05$) post *cis*-aconitate treatment.

Thus, inducing RNAi of the PAD genes was shown to reduce the capacity of pleomorphic cells to differentiate in response to *cis*-aconitate, thus demonstrating that they do play a role in the initiation of differentiation. The residual differentiation observed in these cells was presumably due to incomplete knockdown, or other routes of *cis*-aconitate entry into the cells when exposed to high extracellular concentrations of *cis*-aconitate

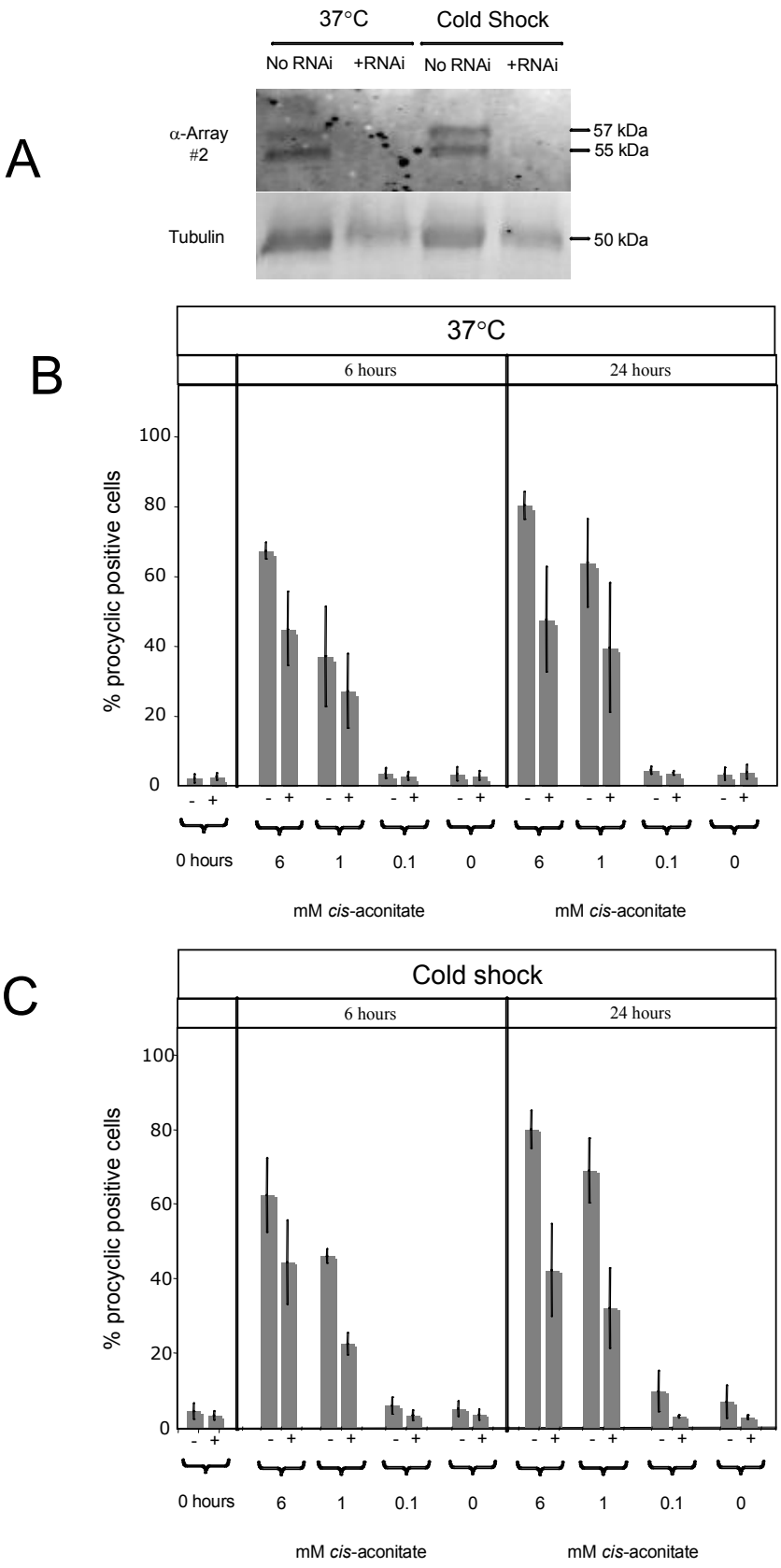

A
37°C
Cold Shock
No RNAi
+RNAi
No RNAi
+RNAi
α-Array #2
57 kDa
55 kDa
Tubulin
50 kDa
B
37°C
6 hours
24 hours
% procyclic positive cells
100
80
60
40
20
0
- +
0 hours
6
1
0.1
0
mM cis-aconitate
C
Cold shock
6 hours
24 hours
% procyclic positive cells
mM cis-aconitate

Figure 6.6-3 Transient RNAi on slender AnTat 1.1 cells. Slender form AnTat 1.1 cells were transiently transfected with an RNAi construct to ablate all the PAD family transcripts. (A) The Western blot shows that RNAi ablated the PAD transcripts such that no PAD proteins were detected in the RNAi cells at 37°C or 20°C. Transfected cells were allowed to recover for 24 hours and subsequently incubated overnight at either (B) 37°C or (C) 20°C prior to the addition of different concentrations of *cis*-aconitate. Thereafter, cells were assayed for differentiation by flow cytometry analysis of EP expression at 0, 6 and 24 hours post *cis*-aconitate addition. Cold shocked RNAi treated cells were significantly impaired in their ability to differentiate in response to *cis*-aconitate at both timepoints ($P < 0.005$), and non-cold shocked RNAi treated cells show statistically significant impairment only at 12 hours ($P < 0.05$). The error bars show the standard error of the data.

6.6.2 Stable RNAi of the PAD array

The data from the transient RNAi showed that ablating PAD transcripts impaired the ability of cells to differentiate in response to *cis*-aconitate. It was therefore of interest to determine whether this effect was recapitulated in a stable, pleomorphic, PAD array RNAi cell line.

To this end, 90:13 AnTaT 1.1 pleomorphic cells containing the T7 polymerase and the Tet repressor protein (a kind gift of Markus Engstler and Michael Boshart) were harvested from an early mouse infection and transfected with the linearised array:pALC14 RNAi construct (see Section 6.3.1). The cells were recovered overnight in methyl cellulose containing media, and transgenic parasites were selected using 0.5 μg ml^{-1} puromycin (see Materials and Methods for details). Eventually, 1 clonal cell line was obtained, which was subsequently inoculated into immunosuppressed mice to generate blood stocks. The generation of this cell line was an important step in determining the function of the PAD proteins.

To determine the affect of PAD mRNA ablation upon the ability of cells to differentiate, 2 immunosuppressed mice were inoculated with the PAD RNAi cell line, and 1 mouse with the 90:13 (parental) cell line. The parental cell line-inoculated mouse, and 1 of the RNAi cell line-inoculated mice, were provided with doxycycline in their drinking water to induce RNAi, and the remaining mouse was provided with normal water. The infections were then monitored for the presence of stumpy forms, and harvested when the population became >90% stumpy in morphology. The cells were subsequently incubated overnight in HMI9 media at either 20°C, or 37°C in the presence or absence of tetracycline as appropriate, such that 6 conditions existed:

- Parental, 37°C +tetracycline (+dox)
- Parental, 20°C +tetracycline (+dox)
- RNAi PAD array, 37°C -tetracycline (-dox)

- RNAi PAD array, 37°C +tetracycline (+dox)
- RNAi PAD array, 20°C -tetracycline (-dox)
- RNAi PAD array, 20°C +tetracycline (+dox)

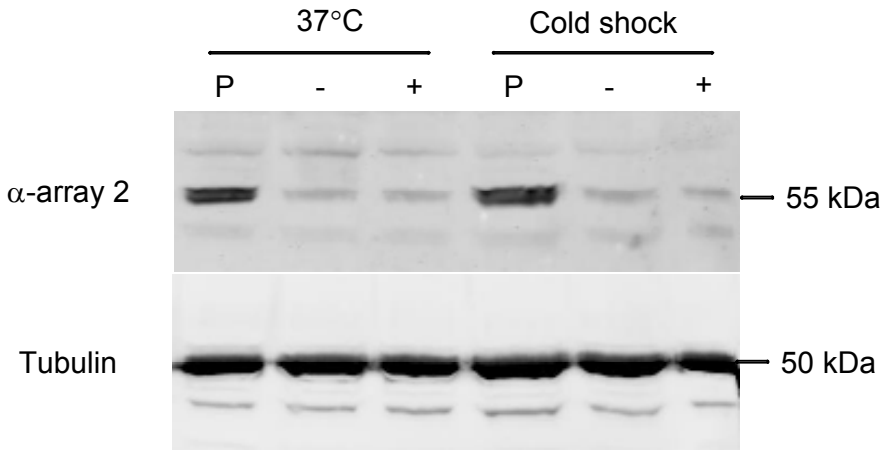
Protein samples were made at this point to determine the extent of PAD protein knockdown. The cells were then treated with differing concentrations of *cis*-aconitate, and differentiation was assayed by flow cytometry analysis of EP procyclin expression at 0, 6 and 24 hours post treatment.

Figure 6.6-5A shows a typical Western blot when samples from 0 hours were analysed using the α-array antibody. It was clear that knockdown had occurred in these cell lines, although did not appear to be clearly inducible, presumably due to leaky RNAi. Analysis of PAD expression using the LI-COR Odyssey system showed that knock down was approximately 6 fold at 37°C, and 7 fold at 20°C with respect to the parental line. The addition of doxycycline did not appear to result in an increased knockdown at 37°C, and only resulted in an increase of approximately 30% ablation in cold shocked cells. Nonetheless, it was clear that the RNAi cell line was reduced in its ability to synthesis PAD array proteins.

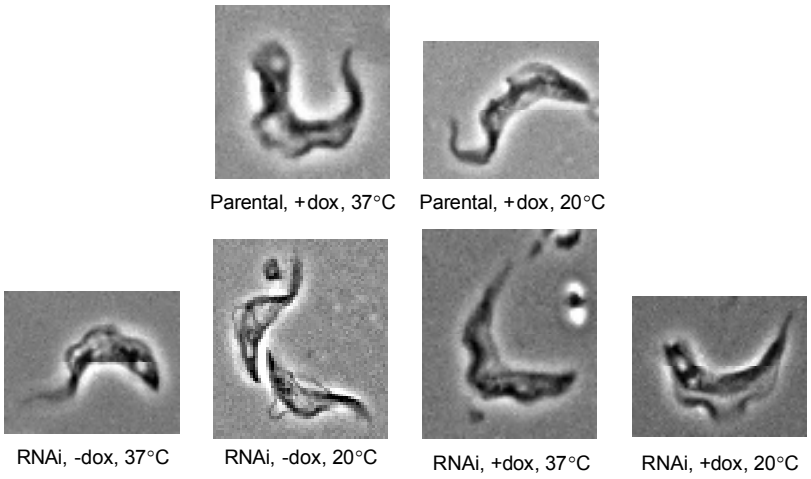
Figure 6.6-5B shows representative cells subsequent to the overnight incubation at 37°C or 20°C. The cells showed clear stumpy morphology, demonstrating that the RNAi cells did not have a reduced pleomorphism, either as a consequence of passaging, or RNAi ablation of the PAD array.

Supporting this, Figure 6.6-5C shows that the 6 populations of cells had similar kinetoplast and nucleus configurations (KN), with high percentages of 1K1N cells, such that each population would be expected to contain similar numbers of stumpy form cells, and would therefore be predicted to differentiate with similar efficiencies unless PAD ablation had an effect.

A



B



C

	1K1N	2K1N	2K2N
Parental 37	93	2	5
Parental CS	96	1	3
Array -dox 37	98	0	2
Array -dox CS	97	1	2
Array +dox 37	96	2	1
Array +dox CS	97	2	1

Figure 6.6-4 The effect of stably inducing RNAi upon pleomorphic cells. A stable pleomorphic PAD array RNAi cell line was generated and inoculated into mice that were given, or not, doxycycline. The cells were allowed to progress to the stumpy form, harvested, and allowed to incubate at 37°C or 20°C overnight in HMI9 containing tetracycline, or not. Parental 90:13 cells were also included, and were inoculated into mice fed with doxycycline. The analysis described here is representative of the 3 repeats of this experiment. (A) A Western blot shows that the RNAi cell line was impaired in its ability to synthesise PAD protein. Analysis using the LI-COR Odyssey system showed that RNAi cells exposed to doxycycline/ tetracycline produced 6 and 7 fold less PAD protein than parental cells at 37°C and 20°C respectively. RNAi was not affected by doxycycline at 37°C, but at 20°C doxycycline treated cells had approximately 30% reduced PAD protein compared untreated RNAi cells. (B) Subsequent to an incubation at 37°C or 20°C, cells were visualised by phase contrast microscopy. All cell lines show a strong stumpy morphology, therefore PAD RNAi did not reduce the capacity for pleomorphism. (C) The different cell line populations were analysed for their nucleus and kinetoplast configuration to confirm that they exhibited similar levels of stumpy form cells.

Figure 6.6-5 shows the data on the capacity for differentiation from 3 repeats of the experiment described above. At 37°C, the RNAi cell lines both showed an inhibition of differentiation at all concentrations of *cis*-aconitate at 6 and 24 hours post differentiation. This effect was subtle, but clear, particularly at lower concentrations of *cis*-aconitate.

The same effect was seen when cells were pre-incubated at 20°C, but the extent was more dramatic and appeared to be inducible, in keeping with the Western blot of PAD protein expression (see Figure 6.6-4A). At 6 mM *cis*-aconitate, cold shocked RNAi cells exposed to tetracycline/ doxycycline differentiated almost 3 fold less efficiently than parental cells at 6 hours. The same effect was observed at 1 mM *cis*-aconitate, and was even more pronounced at 0.1 mM *cis*-aconitate.

A paired t test shows that doxycycline/ tetracycline treated PAD array RNAi cells maintained at 37°C were significantly impaired in their ability to respond to *cis*-aconitate with respect to the parental control cells, both at 6 hours and at 12 hours ($P < 0.005$ for both timepoints). Doxycycline-treated versus non doxycycline-treated RNAi cells, and parental versus non- doxycycline treated cells, do not show significant difference when maintained at 37°C ($P = 0.437$ and 0.101 respectively). Under cold shock conditions, PAD array RNAi cell lines were significantly impaired in their ability to differentiate in response to *cis*-aconitate compared with parental cells, both in the presence and the absence of doxycycline, and at 6 and 24 hours post *cis*-aconitate ($P < 0.0005$). The cold shocked doxycycline-treated RNAi cells were not significantly impaired in their ability to differentiate compared to untreated RNAi cells at 6 hours ($P = 0.095$), but were at 24 hours ($P < 0.05$).

Interestingly, less EP procyclin expression was observed under cold shock conditions at 0 hours, and at 0 mM *cis*-aconitate 6 and 24 hours post differentiation. This had also been seen with the transient assays carried out previously (see Figure 6.6-2), and was thus investigated further.

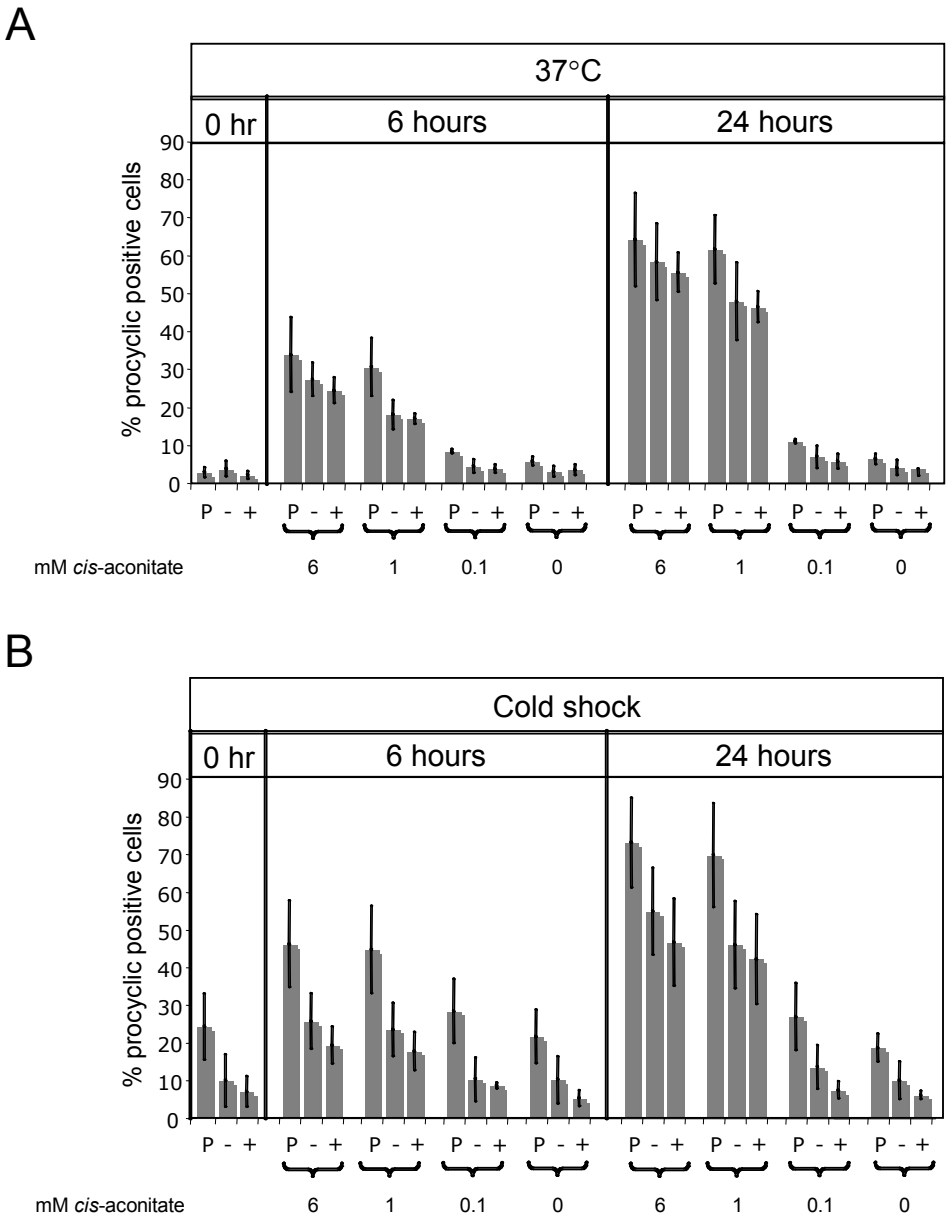


Figure 6.6-5 The effect of stable PAD RNAi upon the differentiation of cells in response to *cis*-aconitate. The average differentiation of RNAi (induced and non-induced with doxycycline) cells and parental cells was assayed by flow cytometry analysis of EP procyclin expression 0, 6 and 24 hours subsequent to the addition of different concentrations of *cis*-aconitate. Cells were pre-incubated at either (A) 37°C or (B) 20°C prior to differentiation induction. Error bars show the standard error of the data.

6.6.3 The effect of PAD RNAi upon cold shock induced EP procyclin upregulation

As discussed, the RNAi cell lines appeared to have reduced cold shock-induced EP procyclin. This suggested that an effect of PAD proteins may be to facilitate expression of EP procyclin in response to cold shock. To dissect this interesting attribute further, the flow cytometry traces of EP expression were compared between the RNAi cell line and the parental cells.

Figure 6.6-6 supports the summary of EP expression shown in Figure 6.6-5B. Thus, the parental cells (red trace) expressed EP procyclin in response to cold shock, showing a proportion of cells highly fluorescent in the FL1 channel (corresponding to EP expression). This matches the findings of Engstler and Boshart (2004). However, the EP induction was greatly reduced in the RNAi cells, and appeared to be inducible, being further reduced in cells treated with doxycycline/ tetracycline.

Thus, it would appear that PAD protein expression is linked to the two attributes of cold shock, (a) hypersensitivity to *cis*-aconitate, and (b) induction of EP procyclin expression.

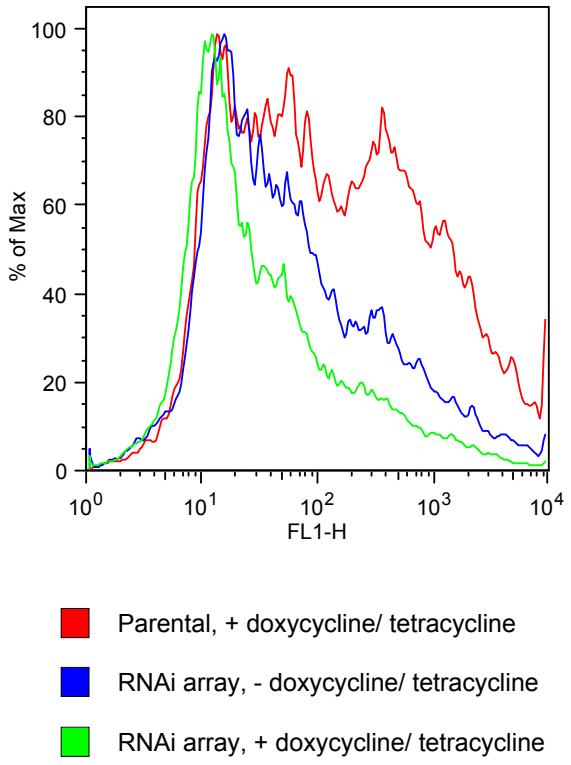


Figure 6.6-6 The effect of PAD ablation upon cold shock induced EP procyclin expression. Parental cells and RNAi cells were grown in mice provided with and without doxycycline treated water. The stumpy form cells were harvested from the mouse, and were incubated overnight at 20°C in HMI9 media, whereby doxycycline exposure was replaced with 1 μg ml^{-1} tetracycline. The extent of EP upregulation was then determined by flow cytometry analysis (FL1 channel). The RNAi cell lines consistently showed a reduced induction of EP in response to cold shock, with the doxycycline/ tetracycline treated cells showing a more pronounced effect than the non-doxycycline/ tetracycline treated cells. Indeed, here the effect was reduced to levels observed when cells were incubated at 37°C.

6.7 Response of the RNAi cell line to an alternative differentiation trigger

The data presented in Sections 6.6.1.1 and 6.6.1.2 show that PAD protein reduction inhibited differentiation in response to *cis*-aconitate. However, this effect was only subtly inducible in response to doxycycline and may represent a cell line defect unrelated to PAD array ablation. This was unlikely because the data from the transient RNAi was consistent with the data from the stable cell line. Nonetheless, it was important to determine whether these cells were inhibited in their ability to differentiate *per se*, and whether the effect was *cis*-aconitate specific.

As discussed (see Section 1.5), a number of differentiation triggers exist in trypanosomes. Pronase was chosen because it has been shown to stimulate differentiation in stumpy form cells with very high efficiency (Hunt *et al*. (1994)) (Sbicego *et al.* (1999)). Therefore, the PAD RNAi cell lines and the parental 90:13 cells were inoculated into immunosuppressed mice and maintained with doxycycline treatment (as described in Section 6.6.2). Stumpy form cells were then harvested and allowed to recover in HMI9 media for 2 hours. Subsequently, the cells were washed in PSG, and treated with, or not, 4 U μl^{-1} of Pronase for 10 minutes at 25°C. Cells were subsequently washed in PSG and re-suspended in SDM79 media at 27°C. Differentiation was then assayed by flow cytometry analysis of EP expression at 6 hours post differentiation.

Figure 6.7-1 shows that the PAD RNAi cell line was not inhibited in its ability to differentiate in response to *cis*-aconitate. Rather, in this experiment it differentiated slightly more efficiently than the parental cells. This probably reflects experimental variation, but it does nonetheless demonstrate that the PAD RNAi cell line is fully able to differentiate when induced by a *cis*-aconitate independent trigger. Hence, PAD proteins mediate *cis*-aconitate dependent signalling.

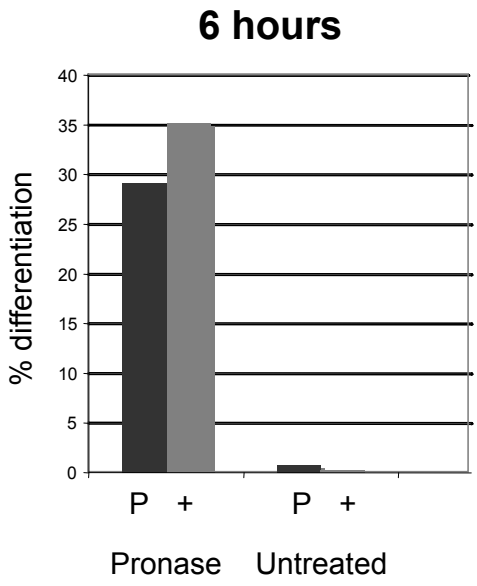


Figure 6.7-1 The ability of PAD RNAi and parental cells to respond to Pronase. To determine whether the reduced differentiation observed in PAD RNAi cells in response to *cis*-aconitate was a cell line specific defect unrelated to the ablation of PAD genes, cells were treated with Pronase, or not, and the differentiation efficiency was observed after 6 hours at 27°C. RNAi cells were not inhibited in their ability to differentiate in response to Pronase. P = 90:13 parental. + = PAD RNAi cells. Both cell lines were harvested from a mouse that was given doxycycline.

6.8 Expression of PAD1 in Xenopus laevis *oocytes to determine citrate transport activity*

The substrate specificity of a transporter molecule can be tested using the *Xenopus laevis* expression system. Synthetic mRNA coding for the protein of interest is microinjected into freshly harvested *Xenopus* oocytes. The oocytes are allowed time to express the protein and localise it to the surface membrane of the cell, and are then assayed for transport activity through incubation with the radio-labelled putative substrate. There are a number of potential problems associated with this strategy:

- the mRNA may not be expressed by the cell
- the protein may not be localised to the cell surface membrane
- the protein may prove toxic to the cell, thus killing the cell before the protein is fully expressed
- the quality of the oocyte must be good in order to achieve reproducible data
- each transporter has different characteristics that require optimisation for the assay to work (e.g. pH, presence of appropriate ions etc)
- some radio-labelled substrates are not commercially available
- the oocyte may contain endogenous transporters that give high background; in addition, endogenous transporters an be upregulated through expression of the exogenous transporter

Despite these issues, this system has been successfully used to identify many transporter substrates, and to investigate the kinetics, and mechanisms of transport (Bröer *et al.* (1997)) (Saliba *et al.* (2006)).

Therefore, PAD1 was cloned into the eukaryotic expression vector pGem-He-Juel (pGHJ) such there was a Kozak sequence adjacent to, and upstream from, the initial ATG start codon in order to enhance ribosome binding, and subsequent protein translation in the oocyte. After sequence verification, a small scale DNA preparation

of pGHJ:PAD1 was used in an *in vitro* mRNA synthesis reaction to generate synthetic PAD1 mRNA (see Materials and Methods). Oocytes were microinjected with the PAD1 cRNA, incubated to allow expression of the protein, and assayed for radio- labelled citrate uptake.

6.8.1 Optimisation of assay

Initial attempts to show PAD1 transport activity suggested that there may be some transport above basal levels, but was not convincing due to high levels of background transport (see Figure 6.8-1A. A number of variables were investigated, including the pH of the transport buffer and the amount of time allowed for the oocytes to express the PAD1 protein. Additionally, the concentration of cold, carrier substrate was varied; cells incubated with no carrier substrate and 5 μM radio-labelled citrate appeared to transport marginally better than those incubated with 5 μM or 50 μM carrier citrate or *cis*-aconitate, suggesting that transport did occur in the μM range. Increasing the pH from pH 7 to pH 9 did not affect the levels of radioactive transport, but did appear to reduce background levels of endogenous transport.

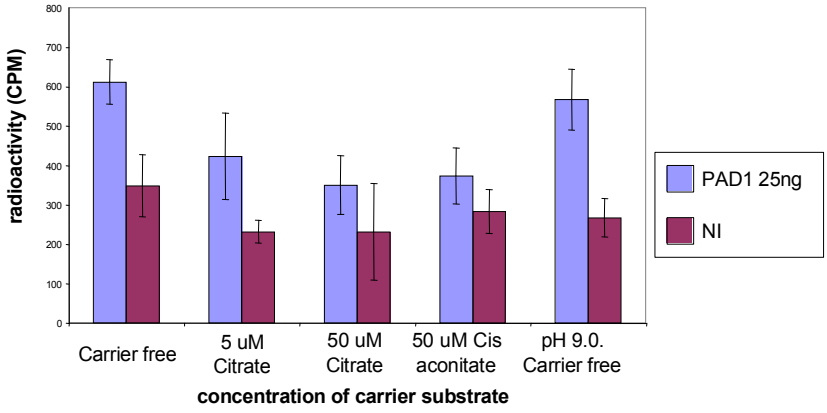
6.8.2 PAD1 is able to transport citrate when expressed in Xenopus laevis

Finally, the quantity of PAD1 cRNA injected into the oocytes was varied. Surprisingly, when the quantity of cRNA injected was titrated from 50 – 5 ng, the levels of PAD1 transport levels rise correspondingly to approximately 3 fold background levels at 5 ng of PAD1 cRNA (see Figure 6.8-1B). An unpaired t test showed that this was statistically significant ($P < 0.0005$). One explanation for this inverse correlation was that excessive intracellular cRNA concentration interfered with endosomal trafficking of transmembrane protein through the endoplasmic reticulum, thus preventing localisation of PAD1 protein to the cell surface membrane.

Further optimisation may increase transport levels further, and competition experiments may help to determine substrate specificity. For example, it would be of interest to determine whether PAD proteins are able to transport *cis*-aconitate (which

is not commercially available as a radio-labelled chemical) as well as citrate. Furthermore, given that *T. brucei* is unable to differentiate in response to isocitrate, it may be revealing to determine whether PAD proteins are also able to transport this tri-carboxylate. Unfortunately these assays could not be completed in the timescale of the project, although they are being pursued in collaboration with Kiaran Kirk's laboratory. Nonetheless, it was clear that the PAD1 protein was able to induce transfer of citrate in *Xenopus laevis* oocytes.

A



B

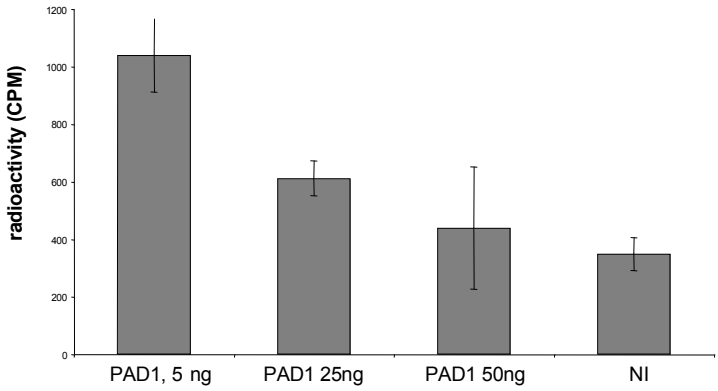


Figure 6.8-1 Transport of 14C citrate in *Xenopus laevis* oocytes micro-injected with PAD1 synthetic mRNA. Synthetic PAD1 mRNA was microinjected into *Xenopus* oocytes and their ability to transport 14C citrate was assayed. (A) Oocytes injected with PAD1 RNA transport 14C citrate at higher levels than non-injected oocytes. Transport levels are low, although above background. (B) 14C citrate transport increased when less PAD1 RNA was injected. Transport of citrate when 5 ng of PAD1 RNA was microinjected was approximately 3 fold above background levels ($P < 0.0005$). Further optimisation may increase this activity.

6.9 Conclusion

The data presented here demonstrates that the PAD genes play an important role in the initiation of differentiation in response to *cis*-aconitate. The hypothesis that the PAD genes promotes differentiation by transporting *cis*-aconitate into the cell is strongly supported. Furthermore, the interesting, and unexpected, observation that PAD ablation reduces cold shock-induced upregulation of EP procyclin adds further support to the role of PAD proteins in *cis*-aconitate mediated differentiation, their link with hypersensitivity to *cis*-aconitate. It would be interesting to dissect in future the mechanism of PAD protein and EP protein co-dependency.

Chapter 7

Discussion

7 Discussion

This study has set out to test the hypothesis that the PAD proteins act as environmental sensors by importing citrate/ *cis*-aconitate, thus enabling cells to differentiate.

7.1 Bioinformatics evidence

Although an important and useful analysis, the bioinformatics data did not provide direct evidence of PAD gene function. The PAD proteins contain both of the motifs characteristic of members of the Major Facilitator Superfamily (MFS) (see Section 3.3.3), and contain 14 predicted *trans* membrane helices (TMS) (see Section 3.6), making these genes members of a little characterised MFS subset, and thus of intrinsic interest. Almost all of the characterised PSI-BLAST hits were carboxylate transporters (see Section 3.8), supporting the hypothesis that PAD genes are *cis*-aconitate/ *cis*-aconitate transporters. Interestingly, the majority of these carboxylate transporters were anti-porters, suggesting that the PAD proteins may facilitate export of one substrate, and the import of a different substrate.

An analysis of the PAD array locus showed that there were several repetitive elements within the intergenic regions (see Section 3.3.3) and allowed the PAD genes to be grouped based on the homology in their 3' UTRs: PAD1/5/7, PAD2, PAD3, PAD4/6/8. This 3' UTR grouping was reflected in mRNA expression, as PAD1, 5 and 7 have similar expression pattern during the parasite lifecycle (see Section 5.1).

7.2 Analysis of the DiD1 defect

As described (see Section 1.8 and Tasker *et al.* (2000)), DiD1 cells do not differentiate in response to citrate/ *cis*-aconitate. Although, the basis of this defect was not resolved by this study, PAD1 and PAD2 were implicated in this phenotype as both transcripts were found to be highly expressed in DiD1 cells with respect to

the parental cell line (see Section 4.1). This was also reflected in protein levels, whereby PAD1 (and probably PAD2) protein was found to be expressed in DiD1 cells, but not in the parental strain (see Section 4.6). This raised a clear disparity; two proteins that were hypothesised to import citrate/ *cis*-aconitate were found to have elevated expression in a cell line that is unable to differentiate in response to *cis*-aconitate. Moreover, no gene sequence differences were found between PAD1 and 2 in the different cell lines (see Section 4.4), and ectopic expression of PAD1 did not inhibit differentiation (see Sections 4.5.1 and 6.1.1), ruling out the possibility of a dominant negative effect.

Therefore, the evidence suggested that the DiD1 phenotype was not due to PAD1 and PAD2 expression *per se*. To investigate the mechanism of PAD1 and PAD2 enrichment in DiD1 cells, a sequence analysis of the PAD1 3'UTR, and a Southern blot analysis of the PAD array locus, was performed, but showed no differences between the DiD1 cells and the parental cells (see Sections 4.3 and 4.2). This demonstrated that the elevated expression of PAD1 and 2 that was observed in the DiD1 cells was not due to a mutated PAD1 3' UTR, nor due to these gene being moved to a more highly transcribed genomic locus. Rather, it was likely that it was due to unidentified *trans* regulatory differences.

As PAD array genes were found to be expressed in stumpy forms (see Section 5.1), other stumpy-enriched transcripts were investigated in DiD1 cells. Of these, PTP1 was the most interesting; it was reported that reduced levels of active PTP1, caused by either RNAi-mediated ablation or pharmacological inhibition, caused blood form cells to spontaneously differentiate to procyclic form cells (Szöor *et al* (2007)). Thus, if elevated levels of PTP were found, this would neatly explain the phenotype. However, none of these genes was found to be aberrantly expressed in DiD1 cells and there did not appear to be a global enrichment of stumpy transcripts (see Section 4.7).

In conclusion, the mechanism underlying the DiD1 defect was not resolved, and it is unlikely that the PAD proteins are directly involved. Therefore, it seems that if

interrogated properly, the DiD1 cell line may have more information to yield. Further work focused establishing the role of the PAD proteins in differentiation.

7.3 Expression of PAD1 in Xenopus laevis *oocytes*

The working hypothesis of this study has rested upon the ability of either PAD1 or PAD2 to transport citrate/ *cis*-aconitate. To investigate this, PAD1 cRNA was micro-injected into *Xenopus laevis* oocytes and tested for enhanced citrate uptake. The data show that PAD1 cRNA injected oocytes were significantly enhanced in their citrate uptake, demonstrating that PAD1 is able to transport citrate (see Section 6.8). The uptake seen was approximately 3 fold that of the control oocytes, and therefore not comparable with the dramatic increases shown by some adenosine transporters (Chiang *et al.* (1999)). This was probably due to the presence of endogenous tri-carboxylate transporters within *Xenopus* oocytes, resulting in high background. Nonetheless, it is likely that with further optimisation transport levels could be increased. Indeed, as potential anti-porters (see Section 3.8), efficient citrate transport may depend upon the presence of a specific, intracellular, substrate.

The possibility remains, however, that the transport observed by PAD1 represents non-specific transport, and that the *in vivo* substrate may not be citrate or *cis*-aconitate. Further work by Rosa Marchetti in Kiaran Kirk's lab (Australian National University, Canberra, Australia) will focus on determining the substrate specificity of the PAD proteins. Nonetheless, these data clearly show that PAD1 is able to transport citrate.

7.4 Expression of PAD genes

As discussed previously (see Section 1.8), macroarray data showed that PAD1 and/ or PAD2 were highly expressed in DiD1 cells with respect to the parental cell line (Pamela Davies, University of Manchester, unpublished data). This study has extended this observation to give a more complete picture of the PAD genes' expression profile.

7.4.1 PAD lifecycle expression and localisation

PAD1 mRNA and protein were shown to be strongly upregulated in stumpy forms, and almost undetectable in monomorphic slender forms and procyclic forms (see Sections 5.1 and 5.4). The identification of a stumpy form-specific protein represents a significant advance as there are currently no molecular markers able to distinguish slender from stumpy forms that are not associated with mitochondrial activation (Matthews and Gull (1994)). Further applications of a stumpy marker are described in Section 7.6.1.

PAD2 mRNA was shown to be expressed predominantly in procyclic forms, but was also detected in stumpy forms (see Section 5.1). Similarly, protein expression of PAD2 increased within 2 hours of differentiation from stumpy to procyclic forms, where it was maximally expressed (see Section 5.5.2 and 5.8). As with PAD1, PAD2 protein was not detected in monomorphic slender forms.

PAD5 and 7 mRNA (which are indistinguishable) was shown to be stumpy form enriched (see Section 5.1). However, a Western blot using a α-array PAD generic antibody did not detect PAD5/7 protein. Instead, two bands of 55 kDa and 57 kDa were detected, but were subsequently shown to be PAD1 and PAD2 respectively (see Section 5.6). This indicated that either PAD5/7 is translationally repressed in stumpy forms (as has been shown with EP isoforms in epimastigotes (Urwyler *et al.* (2005)), or that PAD5/7 proteins resolve at precisely the same size as PAD1 or PAD2, and therefore contribute to either of these bands. This issue could be resolved by either reacting stumpy form protein with a PAD5/7 specific antibody in a Western blot, or by ablating PAD5/7 mRNA in stumpy forms and determining whether the bands detected by the α-array antibody have reduced intensity.

PAD3 was not detected by Northern blot, and PAD4, 6 and 8 were not investigated in this study. Western blots using the α-array PAD generic antibody did not reveal any unaccounted for bands (see Section 5.6), suggesting that PAD family proteins other than PAD1 and 2 are not expressed in slender, stumpy or procyclic forms cells. Interestingly, however, a recent study showed that an RNA binding protein,

TbDRBD3, may regulate PAD4, 6 and 8 transcripts in procyclic forms (Estévez (2008)).

PAD1 was shown to localise to the cell surface membrane of stumpy form cells (see Section 5.2.2.2), which supports the hypothesis that PAD1 promotes differentiation by importing citrate/ *cis*-aconitate into the cell. The localisation and the affect of cold shock upon PAD2 is discussed in Section 7.4.2.

7.4.2 The affect of cold shock upon PAD2

The discovery that citrate/ *cis*-aconitate is able to efficiently induce differentiation to procyclic forms (Brun and Schönenberger (1981)) (see Section 1.5.2) was somewhat perplexing: the differentiation response was specific to these tri-carboxylates, but the high concentration required for this effect excluded them from being an *in vivo* trigger. However, in 2004 Engstler *et al*. reported that when stumpy, but not slender form, cells were subjected to a temperature decrease from 37°C to 20°C, they became sensitive to micromolar concentrations of *cis*-aconitate. The *cis*-aconitate concentration in the tsetse was reported to be 15.9 μM (Hunt *et al*. (1994)), but this analysis use whole tsetse homogenate, rather than determining the concentration in the midgut specifically. However, the concentration of citrate found in the blood that accompanies the trypanosome into the tsetse midgut is approximately 130 μM (Jacobs and Lee (1964)) (Taylor *et al*. (1998)), and is therefore within the range reported to induce differentiation. Thus, citrate/ *cis*-aconitate is a potential physiological trigger of differentiation.

Importantly, when stumpy forms were subjected to cold shock conditions, PAD2 protein was shown to be upregulated 3 to 6 fold (see Section 5.16), with some experiments showing expression of up to 32 fold that of cells maintained at 37°C (see Figure 5.15-3). This raises the intriguing possibility that the upregulation of PAD2 enables hypersensitivity to *cis*-aconitate. However, it could be argued that a 4 fold upregulation of protein expression is unlikely to result in a 1000 fold increase in *cis*-aconitate sensitivity.

There are two responses to this argument. The first is that PAD2 upregulation is unlikely to be the only molecular change that the cell undergoes in order to become hypersensitive to *cis*-aconitate, and there are likely to be a number of intracellular molecules of the *cis*-aconitate reception pathway affected by cold shock. Hence, PAD2 upregulation may be one part of a system optimised for hypersensitivity. The second is that the data presented here show that, in addition to increased levels of PAD2, cold shock causes a re-localisation of PAD2, such that stumpy form cells maintained at 37°C hold PAD2 in intracellular storage vesicles until a temperature drop, whereupon PAD2 is rapidly trafficked to the cell surface membrane (see Section 5.17). The combined effect of the upregulation and the re-localisation would be to vastly increase the number of PAD2 molecules that are present on the cell surface membrane, thus enabling much more efficient import of *cis*-aconitate at low extra-cellular concentrations.

There are examples of similar characteristics shown by other transporters. For example, a mammalian kidney cell aquaporin is held in intracellular vesicles until it is re-localised to the surface membrane in response to a diuretic signalling hormone; this causes an increase in cell permeability to water with much faster kinetics than would be possible if the aquaporin had to be synthesised de novo (Kamsteeg *et al*. (2007)). Importantly, this effect is dependent on the presence of another membrane bound protein that modulates intracellular aquaporin retention (see also model in Figure 7.7-1).

Moreover, a pleomorphic PAD array RNAi cell line was impaired in its ability to express EP procyclin in response to cold shock. It is tempting to speculate that cell surface access of procyclin in blood forms is dependent upon the presence of PAD proteins, which may also explain the flagellar pocket checkpoint control of procyclin surface access in slender form cells (Engstler *et al*. (2004)). There is precedence for this type of mechanism; the cell surface localisation of the monocarboxylate transporters MCT1 and MCT4 is dependent upon another membrane bound protein, CD147, in COS cells (Wilson *et al*. (2005)). To investigate this further, one could: firstly determine whether ablation of procyclin eliminates PAD2 surface access in

response to cold shock and, hence, *cis*-aconitate hypersensitivity and secondly determine whether the cold shock induced procyclin surface expression is prevented in cells where PAD genes are ablated individually.

7.5 Phenotypic analyses

7.5.1 PAD ablation

Initial attempts to generate PAD gene deletion mutants focused on monomorphic cells. Although PAD proteins were not detected in monomorphic cells, this was strategy was pursued because (a) the PAD array was originally identified using the monomorphic DiD1 cell line, (b) monomorphic cells do respond to *cis*-aconitate (albeit only at high concentrations), and may therefore have low levels of PAD proteins undetectable by Western blot, and (c) monomorphic cells are amenable to genetic manipulation. However, these efforts were unsuccessful (see Sections 6.4 and 6.5). Further analysis showed that although gene deletion constructs had integrated into the genome, they had failed to remove the PAD1 gene. This was thought to be a consequence of a genome miss-assembly in this region of the database, differences between the genome reference strain and the 427 cell line that was used in this study or a combination of both.

It is revealing to note that a study of a 28 kb region of chromosome 9 reported that a sub-clone of MITat 1.2 that had been serially passaged showed duplication of a 14.7 kb genomic fragment in a region containing tandemly repeated genes (Kramer *et al*. (2007)) Additionally, another study to compare the chromosome 1 sequence of the genome reference strain with Lister 427 showed that, in addition to subtelomeric VSG gene duplications, all arrays of tandemly repeated sequences showed significant copy number differences between the 927 reference strain, and the Lister 427 cell line (Callejas *et al.* (2006)). Thus, irrespective of technical difficulties associated with gene knockouts, it is highly likely that there are PAD gene copy number differences between different cell lines of *T. brucei*, making targeted gene deletion difficult.

Therefore, further work focused on pleomorphic cells. This had the advantage that both PAD1 and PAD2 protein were highly expressed in the stumpy form, and that these cells differentiate synchronously in response to *cis*-aconitate. However, transgenic studies using pleomorphic cell lines have been largely unsuccessful, and very few groups have successfully achieved this. This is probably due to the combined effect of two principal factors. Firstly, pleomorphic cells differentiate to the terminally arrested stumpy form in response to SIF (Vassella *et al.* (1997)); thus if cells are not closely monitored in culture then they enter cell cycle arrest and subsequently die. Secondly, as with other trypanosome isolates, a period of adaptation to culture conditions may be required before successful growth *in vitro*. Despite these caveats, it was clear that it was not going to be possible to obtain useful data on PAD function without using pleomorphic cells.

Prior to the successful cloning of a stable, pleomorphic PAD array RNAi cell line, a transient transfection was used to abate all 8 PAD genes in pleomorphic cells. This strategy had the advantage that it (a) did not require accurate genome sequence covering the PAD array locus on chromosome 7, (b) avoided the technical issues associated with generating stable transgenic pleomorphic cells described above, and (c) avoided difficulties of inducibility, the effects of parasitaemia stage and the inherent possibility of a founder effect associated with cloning cell lines. This strategy ultimately gave reproducible and statistically significant inhibition of differentiation in response to *cis*-aconitate.

To verify this result in stable RNAi cell lines, we made use of a pleomorphic cell line that expressed the tetracycline repressor protein and the T7 polymerase (a kind gift of Markus Engstler and Michael Boshart). In accordance with the transient RNAi data, differentiation was shown to be clearly inhibited when PAD array gene expression was silenced. Interestingly, the most profound inhibition was observed at low concentrations of *cis*-aconitate, at which differentiation was completely abrogated (see Section 6.2.2). This has the implication that, at high concentrations, *cis*-aconitate is transported into the cell via other, lower affinity, transporters, or possibly via non-carrier mediated diffusion across the cell surface membrane. This would explain the efficient differentiation of monomorphic cells at 6 mM *cis*-aconitate in

the absence of detectable PAD array protein expression (for further discussion on this see Section 7.7).

There are two potential caveats associated with this phenotype. Firstly, PAD gene ablation was associated with reduced procyclin expression in response to cold shock (see Sections 6.6.3 and 7.4.2), and this could account for the differentiation phenotype. However, differentiation was also reduced when cells were not previously cold shocked (see Section 6.6.2). Further analysis of differentiated cells could resolve this; for example, analysis of the cells with other procyclic specific markers, such as CAP5.5.

Secondly, PAD2 was shown to be expressed in procyclic forms; if PAD2 is essential in these forms then the differentiation phenotype could be due to non-viability of these cells. This second caveat is unlikely, however, as the timepoints taken for the differentiation assay were too early (6 hours and 24 hours post-differentiation) for an affect on cell growth to complicate the analysis (Brun and Schönenberger (1981)) (Ziegelbauer *et al*. (1990)) (Matthews and Gull (1994)).

Nonetheless, both of these caveats were addressed when these RNAi cell lines were shown to differentiate as efficiently as parental cells in response to proteolytic stress (see Section 6.7). Moreover, this assay also addressed the concern that the stable RNAi cell line had a reduced capacity for differentiation unrelated to PAD gene ablation. Interestingly, this suggests that *cis*-aconitate operates via a different pathway to proteolytic stress induced differentiation.

Thus, it would appear that PAD gene expression is necessary for the initiation of differentiation in response to *cis*-aconitate under physiologically relevant conditions.

7.5.2 PAD ectopic expression

As discussed (see Section 7.5.1), the PAD proteins appeared to be necessary for hypersensitivity to *cis*-aconitate in stumpy form cells. As monomorphic slender cells did not appear to express PAD1 protein (see Section 5.4), it was of interest to determine whether ectopic expression of PAD1 would increase their sensitivity to

cis-aconitate. Further, if *cis*-aconitate import was the limiting factor in the initiation of differentiation, one would expect an increase in differentiation capacity, i.e. at a given *cis*-aconitate concentration, cells expressing PAD1 would be expected to initiate differentiation at a higher frequency than those that do not. Ectopic expression of PAD1 at the cell surface membrane was achieved, albeit at levels lower than on stumpy form cells (see Section 6.1). However, neither of these effects was clearly observed, even under cold shock conditions (see Section 6.1).

It is difficult to draw many firm conclusions from these observations. One would expect that cold shock stimulates a number of molecular changes in stumpy forms, including parts of the intracellular signalling pathway associated with *cis*-aconitate reception. These stumpy attributes or pathways are unlikely to be functional in the serially passaged monomorphic Lister 427 cell line. Additionally, it should be noted that ectopic expression experiments were performed only for PAD1, whereas it was PAD2 that was most clearly linked with hypersensitivity (i.e. PAD2 has increased cell surface membrane expression in response to cold shock). This was unavoidable as these experiments were performed prior to any of the analyses on PAD2.

Nonetheless, it would appear that simply expressing PAD1 was not sufficient to either increase differentiation efficiency or sensitivity to *cis*-aconitate in monomorphic cells. It would be interesting to repeat these ectopic expression experiments in pleomorphic cells to determine whether PAD array expression is both necessary and sufficient for hypersensitivity to *cis*-aconitate.

7.6 Future directions

7.6.1 Applications of a stumpy specific protein

The identification of a stumpy specific protein will be a valuable tool in dissecting the biology of trypanosome transmission. Currently, the only stumpy specific characteristics that have been identified are morphology and mitochondrial activation. Both have been successfully used (Matthews and Gull (1994)) (Matthews (1999)), but both have their limitations. Specifically, morphology can be very ambiguous in a pleomorphic population that contains intermediate forms, and

mitochondrial activation is associated with oxidative stress and cell death in eukaryotes (Giffin *et al.* (1986)). (McCann *et al*. (1986)) (Giffin *et al.* (1989)). Hence, it would be interesting and informative to use PAD1 as a marker for stumpy forms and to correlate this with other stumpy form characteristics. For example, the presence of low numbers of dividing, PAD1 positive cells suggests that commitment to the stumpy form often occurs prior to the final 'differentiation-division', and precedes morphological change. A cytological marker for stumpy forms may yield new insights into the stumpy transformation, and its underlying molecular events.

A further use of a stumpy form marker would be to generate a bioassay for stumpy formation. For example, expressing a reporter gene fused to the PAD1 3' UTR in pleomorphic cells would allow the rapid and objective quantitation of stumpy form cells within the population. This would allow the population dynamics of the slender to stumpy transition to be analysed in a natural infection, potentially providing new insights into trypanosome biology, and the interaction of different cell forms with the mammalian immune system. The development of a reporter cell line will therefore be addressed by Paula MacGregor, University of Edinburgh.

A reproducible and robust bioassay would allow a large number of compounds to be screened for stumpy induction activity, thus facilitating purification of the, so far, unidentified SIF molecule (Reuner *et al.* (1997)) (Vassella *et al.* (1997)). Previous work has been hindered by the absence of a quantitative stumpy bioassay (Saville and Seed (2004)) (MacGregor and Matthews (2008)).

As described (see Section 1.4), a number of chemicals have been reported to bypass the normal route of SIF signalling, and induce slender to stumpy formation (Vassella *et al*. (1997)), even in monomorphic slender cells (Giffin *et al.* (1986)) (Laxman *et al*. (2006)) (Denninger *et al.* (2007)). A stumpy specific marker and a reporter cell line would enable validation of these chemicals.

7.6.2 Further dissection of the role of PAD proteins in promoting differentiation

It would be of further interest to dissect the PAD proteins' relative contribution to differentiation by ablating each gene individually. Initial data on this using transient

transfection was encouraging (data not shown), but further work could make use of existing pleomorphic cell lines that have RNAi constructs stably integrated that target the genes individually. There was insufficient time to analyse these cells in this study.

All of the work in this study has been performed *in vitro*. It would therefore be of great interest to repeat some of these analyses *in vivo*, and determine the ability of PAD RNAi stumpy forms to generate tsetse midgut infections. This would potentiate the quantitative study of the contribution of *cis*-aconitate mediated differentiation with respect to some of the other triggers of differentiation that have been identified. However, it is worth noting that DiD1 cells, although unable to respond to *cis*-aconitate *in vitro*, were able to establish a tsetse midgut infection (Maria Tasker, University of Manchester, unpublished observations).

7.6.3 The roles of PAD family members in other lifecycle stages

The PAD family consists of 8 members, 2 of which have been characterised in some detail, and the mRNA expression profile of 2 more is also known. Thus, the expression and function of other PAD members is also of interest. The expression profile of PAD4, 6 and 8 was not investigated, and the expression of any of the genes in the mesocyclic, epimastigote or metacyclic forms was not tested. Staining cells from these lifecycle stages with the α-array PAD generic antibody could be informative. Analysis of mRNA transcripts using RT-PCR from these lifecycle stages has been described (Urwyler *et al.* (2005)) and could be applied to the PAD transcripts.

7.6.4 Further analysis of the *cis*-aconitate mediated differentiation pathway

It has become increasingly clear that temperature reduction to 20°C is an integral part of *cis*-aconitate mediated differentiation *in vivo* (Engstler *et al.* (2004)). Further work could ablate EP procyclin transcripts to determine whether this inhibits (a) hypersensitivity to *cis*-aconitate and (b) upregulation and re-localisation of PAD2. Furthermore, a mutational analysis of PAD2 could help to identify sequences important for re-localisation, and thus elucidate this interesting phenomenon.

The data presented here suggest that the PAD proteins have an important role to play in the hypersensitivity to *cis*-aconitate; however, they are very unlikely to be the only cellular factors affected by cold shock. A proteomic analysis of cold shocked versus non-cold shocked cells may help to identify other, intracellular, factors that are part of the hypersensitivity system. A microarray may also be revealing, although cold shock may not operate via regulation of mRNA transcripts.

7.6.5 Citrate/ *cis*-aconitate uptake assays

Previous work (Hunt *et al*. (1994)) reported that blood form cells do not import significant levels of citrate. Given the putative role of the PAD genes this is somewhat surprising, and it would be interesting to determine whether cold shocked stumpy forms have increased citrate import activity. This analysis could be extended to include *cis*-aconitate and to determine whether tri-carboxylate import is reduced in the PAD RNAi cell lines.

Furthermore, an analysis to mutate charged residues in the PAD proteins' TMS could determine which residues are important for substrate binding and substrate translocation.

7.7 Insights into trypanosome biology from this work

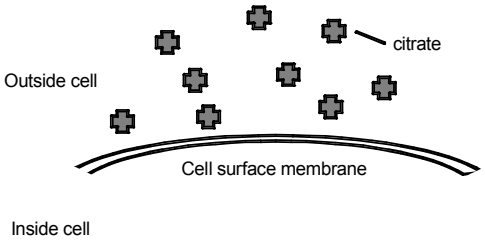
This study has convincingly shown that *cis*-aconitate induced differentiation under physiologically relevant conditions is dependant on the presence of the PAD family of transporters. This has a number of interesting implications. The first is that *cis*-aconitate acts intracellularly, and does not bind to a cell surface receptor in order to exert its effect. This is an unexpected addition to the field as it has long been supposed to exert its effect outside the cell. The findings of Hunt (Hunt *et al.* (1994)) have generally been interpreted to mean that blood form trypanosomes do not transport citrate. The authors showed that pleomorphic trypanosomes exposed to 3 mM extracellular citrate had undetectable levels of intracellular citrate within the first 60 minutes of differentiation, approximately 18 μM after 72 hours, and to 50 μM in established procyclic forms. However, it was not clear from that study whether the cells were taken from a late stage parasitaemia (and hence contained a

stumpy form population expressing PAD family proteins). Importantly, this study was performed prior to the work of Engstler *et al.* (2004), and uptake experiments were therefore not performed under cold shock conditions, i.e. those reported to induce hypersensitivity to *cis*-aconitate.

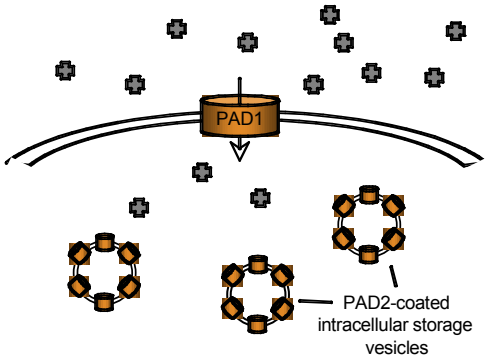
Monomorphic slender cells are capable of differentiating in response to high levels of *cis*-aconitate. The data presented in this study suggest that this may in fact be due to non-specific tri-carboxylate uptake, either through non-specific carrier mediated transport, or diffusion though the cell membrane. However, expression of PAD1 to the cell surface membrane does not appear to increase transformation efficiency or sensitivity (see Section 6.1). It was therefore speculated that perhaps an intracellular molecule is also affected by cold shock; however, cold shock of these cells in combination with PAD1 expression did not enhance differentiation (see Section 6.1). This supports the hypothesis proposed by Engstler *et al*. that stumpy form specific factors are affected by cold shock and enable hypersensitivity to *cis*-aconitate. Identification of one of these factors is described in this study (i.e. PAD2 expression and localisation), clearly there are other, intracellular factors still to be identified.

Our understanding of trypanosome lifecycle biology has advanced much in the last 4 years. The observations regarding the affect of cold shock upon sensitivity to *cis*-aconitate reported by Engstler *et al*. (2004) has meant that this chemical has become a physiologically relevant candidate for stimulating transformation, rather than simply a useful tool. The subsequent discovery of a PTP phosphatase that acts as a break upon differentiation when active (Szöor *et al.* (2007)) and a putative PTP1 interacting protein (currently termed PTPInt) that is cold shock induced, and may have citrate specificity (Szöor, unpublished data), allows us to propose a model of tri-carboxylate induced differentiation into which the PAD story fits neatly (see Figure 7.7-1). Hence, stumpy form trypanosomes are taken up into the tsetse upon feeding, and are subsequently exposed to a low temperature of 20°C. This causes the upregulation and re-localisation of PAD2, which, in concert with PAD1, allows the transport of very low levels of citrate and *cis*-aconitate that are present in the blood meal, and possibly the tsetse midgut. Subsequent binding of citrate to the cold

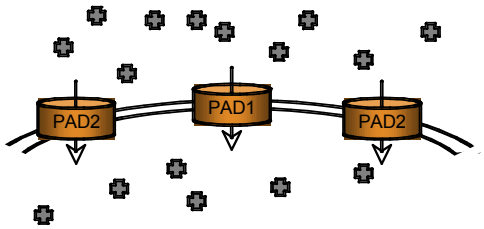
shock-induced PTPInt causes inactivation of PTP1, thus allowing differentiation to proceed, assisted by the activity of gene regulators, such as the CCCH zinc finger family of proteins (Paterou *et al*. (2006)) (Walrad, unpublished data). Clearly, much of this model is speculation, and some parts have more support from the data than others. Nonetheless, the field has progressed to a point whereby these models can be proposed and therefore tested experimentally.



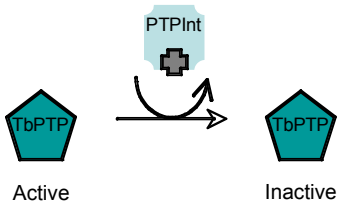
Slender
PAD1 and PAD2 are not expressed in slender forms, therefore citrate cannot enter the cell at physiologically relevant concentrations (~0.1mM).



Stumpy, 37°C
PAD1 is expressed on the cell surface of stumpy forms, whereas PAD2 is held in intracellular vesicles. Differentiation can only occur at very high citrate/ *cis*-aconitate concentrations (1mM)



Stumpy, 20°C
PAD2 protein is upregulated, and expressed on the cell surface. Citrate/ *cis*-aconitate is transported into the cell with high efficiency even at very low extra-cellular concentrations. Intracellular signaling proteins (such as PTPInt) may also be upregulated in response to cold shock.



Down-stream signaling pathway
Citrate/ *cis*-aconitate binds to PTPInt, allowing PTPInt to inactivate TbPTP. TbPTP no longer acts as a break upon differentiation, and stumpy forms initiate differentiation to the procyclic form.

Figure 7.7-1 A model of citrate/ *cis*-aconitate mediated differentiation

Appendix A

Northern blot solutions

[10x] MOPS

MOPS	46.26 g/l
Na-Acetate (pH 7.0)	50 mM
EDTA	10 mM

Autoclaved and stored in the dark at 4 °C. Made up to 1× with dH2O

RNA Gel Loading Buffer

Formamide	30 %
Formaldehyde	6.1 %
Bromophenol blue	0.01 %
MOPS	1 ×
glycerol	10 %

Maleic Acid Buffer

Maleic acid	100mM
NaCl	150mM

pH7.5 with NaOH, autoclaved

10% Block Solution

Blocking reagent	10 % (w/v)
Maleic Acid	1 ×

Wash Buffer

Maleic Acid Buffer	1 ×
Tween 20	0.3 %

Detection Buffer

Tris-HCl	100 mM
NaCl	100 mM

pH 9.5

Southern blot solutions

TELT	
TrisCl, pH 8.0	50 mM
EDTA, pH 9.0	62.5 mM
LiCl	2.5 M
Triton X-100	4% (v/v)
Denaturing Solution	
NaCl	1.5 M
NaOH	0.5 M
Neutralizing solution	
NaCl	1.5 M
Tris	1 M
pH to 7.4 with HCl	
20xSSC	
NaCl	3 M
Tri Sodium Citrate	0.3 M
Hybridization/ pre-hybridization buffer	
Dextran sulphate (dissolved in water first)	5 %
SSC	5 x
Liquid block	10 %
SDS (added last)	0.1%
Buffer A	
Tris pH9.5	100 mM
NaCl	300 mM
Blocking buffer	
Liquid block	10%
Tris pH9.5	100 mM
NaCl	300 mM

Western blot solutions

[10x] Running Buffer	
Tris-HCl, pH8.3	0.25 M
Glycine	1.92 M
SDS	1%
Laemmli Sample Buffer	
Tris-HCl, pH6.8	62.5 mM
SDS	2 %
Glycerol	10 %
Bromophenol blue	20 μg
Coomassie Blue Stain	
Methanol	50%
Acetic Acid	10%
Coomassie Brilliant blue R250	0.1 % w/v
Destain Solution	
Methanol	40%
Acetic Acid	10%
Water	50%
[10x] Blot Transfer Buffer Stock	
Tris (do not pH)	0.025 M
Glycine	0.15 M
1× Completed Transfer Buffer	
SDS	0.2 %
[10x] Transfer Buffer Stock	10 %
Methanol	20 %
Ponceau Stain	
Ponceau S	0.4%
TCA	3%

Immunofluorescence solutions

DAPI working stock

4',6-diamidino-2-phenylindole (DAPI)	10 μg ml-1

MOWIOL Mounting Medium

Glycerol	25 % w/v
MOWIOL	10 % w/v
Tris pH8.5	0.1 M

Medium required heating at 50 °C to dissolve MOWIOL. Aliquots were stored at -20 °C.

Voorheis's modified PBS (vPBS)

NaCl	137 mM
KCl	3 mM
Na2HPO4	16 mM
KH2PO4	3 mM
Sucrose	46 mM
Glucose	10 mM
pH7.6	

Phosphate Buffer Solution (PBS)

NaCl	137 mM
KCl	3 mM
Na2HPO4	16 mM
KH2PO4	3 mM
pH7.6	

Miscellaneous solutions

TAE 1×	
Tris-Acetate	40 mM
EDTA	1 mM
pH 8.0	
TBS 10x	
NaCl	140mM
Tris	10mM
ND96	
NaCl	96 mM
KCl	2 mM
CaCl2	1.8 mM
MgCl2	1 mM
HEPES	5 mM
pH 7.4	
TE	
TrisCl	10 mM
EDTA	1 mM
pH 8.0	

Small scale plasmid preperation solutions

Solution I

glucose	50 mM
EDTA	10 mM
Tris-HC1 (pH 8.0)	25 mM

Solution II

NaOH	0.2 M
Sodium dodecyl sulfate (SDS).	1%

Solution III

Potassium acetate	3 M

pH 5.2

Antibody concentrations

Immunofluorescence

α-EP procyclin	1:500
α-PAD1 antibodies	1:100
α-PAD2 antibodies	1:100
α-mouse-ALEXA 488	1:500
α-mouse-FITC	1:500
α-mouse-TRITC	1:500
α-rabbit-FITC	1:500
α-rabbit-cy3	1:500

Western blot

α-PAD1 antibodies	1:1000
α-PAD2 antibodies	1:1000
α-array antibodies	1:1000
α-mouse-HRP	1:5000
α-rabbit-HRP	1:5000
IRDye 800CW Goat Anti-Rabbit IgG (LI-COR, red)	1:5000
IRDye 680 Goat Anti-Mouse IgG (LI-COR, green)	1:15000

Flow cytometry

α-EP procyclin	1:500
α-PAD1	1:500
α-mouse-FITC	1:500
α-rabbit-FITC	1:500

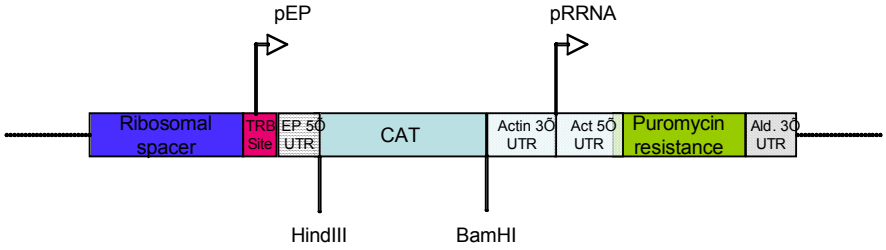
A list of primers used in this study

Primer Number	*Primer name*	*Primer Sequence*	*Comment*
		Sequencing PAD genes	
1	Seq PAD1 FOR1	CCACCATCACTACTGTTGGTATCGC	Binds to PAD1/2/5/7
2	Seq PaD1 FOR2	GTTTCCATCCAACCGTGGTGCCG	Binds to PAD1/2/8
3	Seq PaD1 FOR3	CAAGGGTGTGTACCTGAAGCAGAAG	Binds to PAD1/2/4/5/7
4	Seq PaD1 FOR4	GTTGAAGCGGATGATGAAGTTTCCG	Binds to PAD2
5	Seq PaD1 FOR5	CAACGGTGTGGGTAGTGCTGTAGG	Binds to PAD1/2
6	Seq PaD1 FOR6	CCAAGGATCCTGCAAAGCATTACAAC	Binds to PAD1/2/3/4/6/8
7	Seq PaD1 FOR7	CGACGGAGGATAAAGGAAGCTGAAG	Binds to PAD2
8	Seq PaD1 REV1	GGGAAGGCGAGACCTAACATAAATG	Binds to PAD1/2/4/8
9	Seq PaD1 REV2	GATGAAGTATGGTAAAGGTAATGCAGC	Binds to PAD1/4/5/8
10	Seq PaD1 REV3	CGAGCATTGTAGATGATAACATAGTG	Binds to PAD1+2
11	Seq PaD1 REV5	GAGGGTACCAACGACCAGTGAAAAC	Binds to PAD1/2
12	Seq PaD1 REV6	CAGATCCTTCAATCACTTCTTGGAATG	Binds to PAD1/2/4/5/6/8
13	Seq PaD1 REV4	CAATAACTGTCACAGCAAACCCCAC	Binds to PAD1/2/3/5
14	Seq PaD1 REV7	GCAGAAGGTTCCGAGAACCAAGAG	Binds to PAD2
		Cloning into the *Xenopus* expression vector pGHJ	
77	PAD FOR (EcoRI)	AAAGAATTCCCCGCCACCATGAGCGCACCCGTCGACAACGTC	Amplifies PAD1 open reading frame for cloning into pGHJ
78	PAD1 REV (HindIII)	ATAAAGCTTTTCATTGCGGAGCAGCCTCACGGG	
		Inserting a XhoI site into pHD451	
24	PHD451 mod FOR	AGCTTGAATTCCTCGAGG	When annealed, generates a duplex containing XhoI
25	PHD451 mod REV	GATCCCTCGAGGAATTCA	
		Primers for cloning PAD genes into expression vectors	
26	Gene 1 FOR - PHD451 (HindIII)	TAAAAGCTTATGAGCGCACCCGTCGACAACG	A generic PAD forward primer

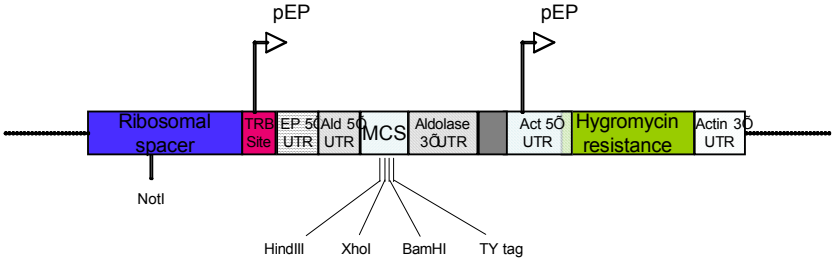
27	Gene 1 REV - PHD451 (XhoI)	TAACTCGAGTTGCGGAGCAGCCTCACGGGC	For cloning PAD1 into pHD451-TY
28	Gene 1 REV - STOP (XhoI)	TAACTCGAGTCATTGCGGAGCAGCCTCACGG	For cloning PAD1 into pHD451 (no tag)
		Primers for generating RNAi constructs	
108	108:pLEW100x mid-PAD1 (HindIII+BclI)	tttaagctttgatcaCCGGGAGGGCGAACGTGGACC	Generates a 76 nucleotide amplicon for cloning into pALC14 and ablating PAD1
109	109: pLEW100x PAD1 REV (XhoI+NdeI)	aaactcgagcatatgTCATTGCGGAGCAGCCTCACGGGC	
106	106: pMP6 PAD1 FOR (HindIII-BclI):	tttaagctttgatcaATGAGCGCACCCGTCGACAACGTC	Amplifies the entire PAD1 open reading frame for cloning into pALC14 (in combination with #109)
		Primers for making knockout constructs	
67	67: A FOR (KpnI)	CCCGGTACCGTAACGTTCAAAAGTCAGCGACAGG	Generates target sequence A
68	68: A REV (XhoI)	AAActcgagAAAATTTATTGACTCCGAGAG	
46	5' array hyg (SacI) KO FOR	CCCGAGCTCCTCACAAAACGCCGCAAGTAATAAAG	Generates target sequence B
45	5' gene 1 inner KO (NotI) REV	AAAGCGGCCGCGTTAGTCTGTGGTATTGTTCAAAGGTTC	
69	69: C FOR (NotI)	TTTGCGGCCGCCTGTTTTAAGACAAGGCAATTGAGAAG	Generates target sequence C
70	70: C REV (SacI)	CCCGAGCTCCCTGCTGTCAACCATTCACCTGCC	
61	61: D FOR (XhoI):	TTTCTCGAGGTTGGAAGGTTTGTTGTTCATGCGTTC	Generates target sequence D
62	62: D REV (KpnI):	CCCGGTACCACCGTACACTGCAGTCAACGCC	
71	71: E FOR (NotI)	TTTGCGGCCGCGCTCTCGTTTATGATCGATAAAAGTTCG	Generates target sequence E
72	72: E REV (SacI)	CCCGAGCTCGGGCGGATATAATGCTATTGATGTGTG	
65	65: F FOR (XhoI):	TTTCTCGAGCGAGGGATGCCCGTGAGGCTGC	Generates target sequence F
66	66: F REV (KpnI):	CCCGGTACCGTGAAGAATTTAACGATATAAGTTGGTGTG	
75	75: PAD array 5' FOR (KpnI)	CCCGGTACCccgatattggtgagttagacaggtg	Generates target sequence G (in combination with #68)
76	76: PAD array 3' REV (SacI):	CCCGAGCTCaagatgccttaggcgcgacctgc	Generates target sequence H (in combination with #69)
		Replacing pHD617 Hygromycin cassette with puromycin cassette	

92	92: EP Procyclin 5'UTR FOR	GGTTCGCTCGAGTTTCCACCCAGC
93	93: EP procyclin 5'UTR REV NdeI	attcatatgcTCTGAGAGCGGTCAGTTGCTCTC
94	94: Actin 5'UTR pHD1034 FOR NdeI	taacatatgCACAGCAAGGTCTTCTGAAATTCATG
95	95: Backbone pHD1034 REV AvrII RC	aatcctaggGATGACCCTGCTGATTGGTTCGCTG

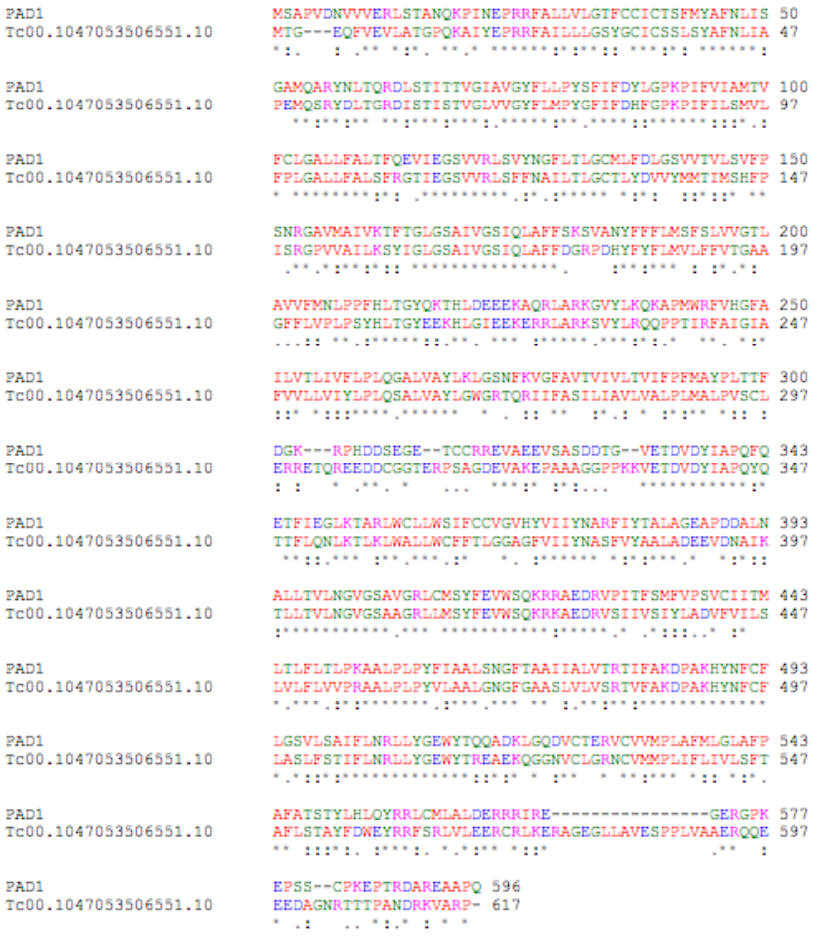
Plasmid maps



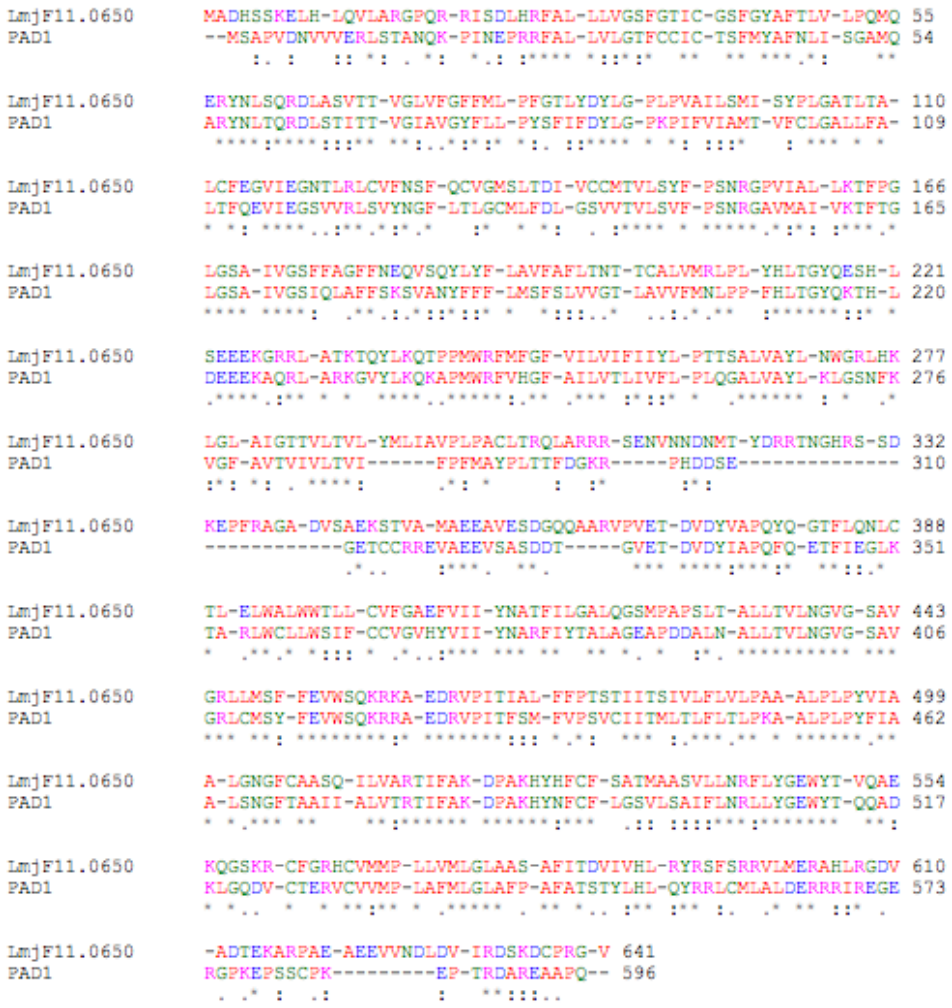
pHD617(puroR)
6.25 kb



pHD451-TY (XhoI)
5.8 kb



An alignment of PAD1 with the closest ortholog in *Trypanosoma cruzi*.

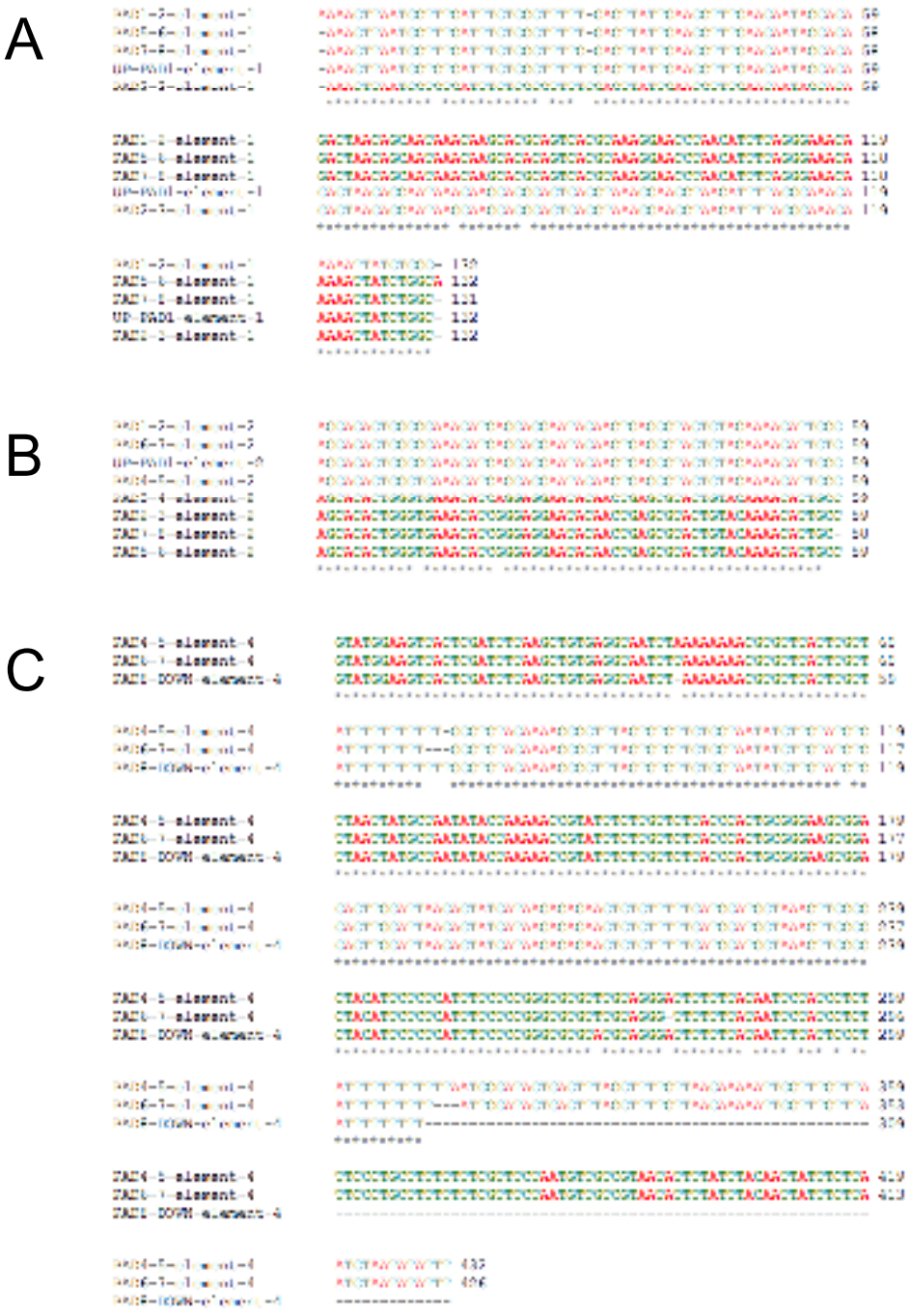


A clustalW alignment of PAD1 with the closest ortholog in *Leishmania major*.

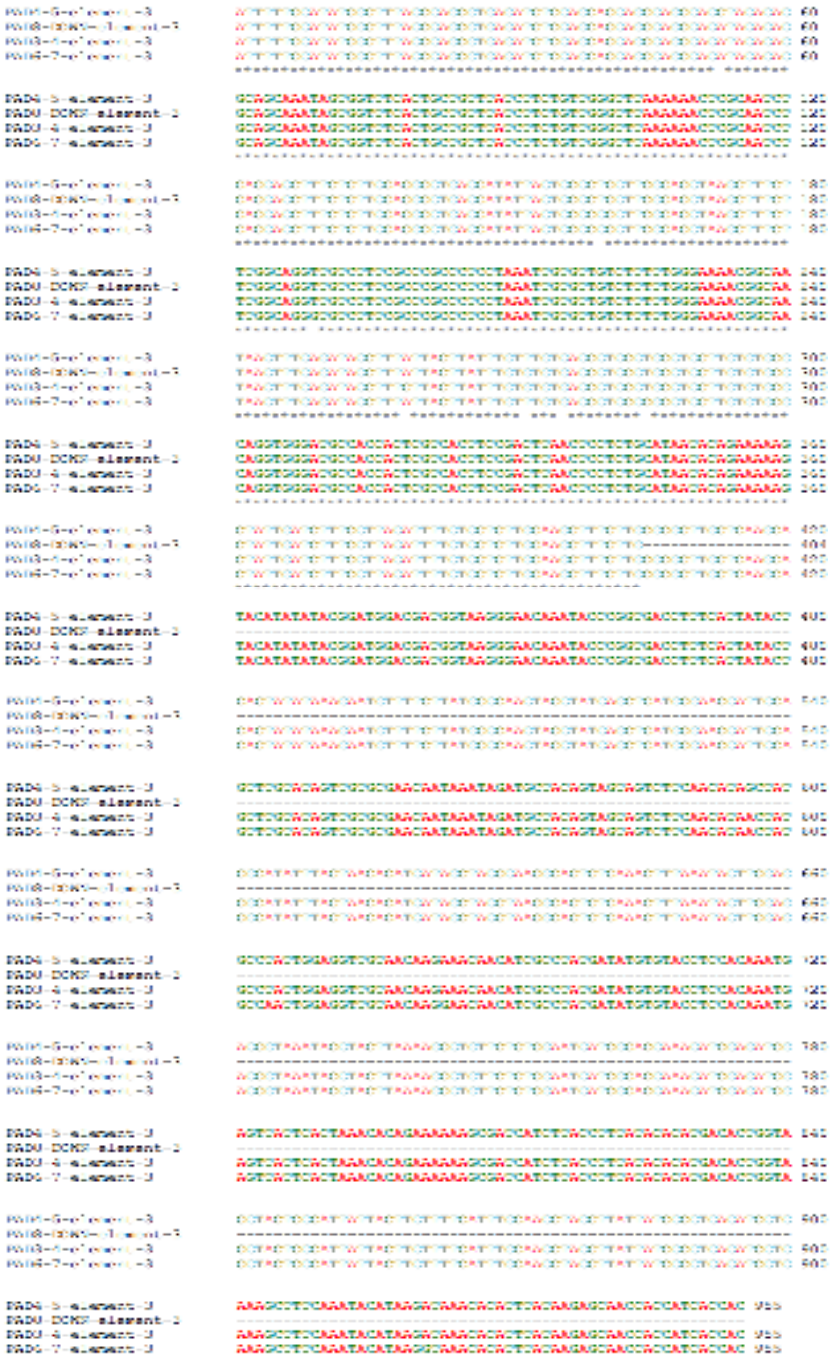
A list of *T. cruzi* PAD1 orthologs.

PAD1 was used in a TBLASTX search against the *Trypanosoma cruzi* genome. The closest matches are shown.

Gene Accession number	Gene description	E value
Tc00.1047053506551.10	Hypothetical protein, conserved|.	4.70E-168
Tc00.1047053508799.270	Hypothetical protein, conserved.	1.30E-167
Tc00.1047053510069.20	Hypothetical protein, conserved|.	1.80E-166
Tc00.1047053509713.30	Hypothetical protein, conserved|.	6.40E-156
Tc00.1047053510103.24	Hypothetical protein, conserved|.	2.10E-155
Tc00.1047053510811.20	Hypothetical protein, conserved|.	5.00E-154
Tc00.1047053508799.260	Hypothetical protein, conserved.	1.10E-151
Tc00.1047053509707.10	Hypothetical protein, conserved|.	2.20E-151
Tc00.1047053508065.70	Hypothetical protein, conserved|.	2.80E-151
Tc00.1047053509713.20	Hypothetical protein, conserved|.	3.30E-147
Tc00.1047053510811.10	Hypothetical protein, conserved|.	9.50E-147
Tc00.1047053509713.10	Hypothetical protein, conserved|.	2.90E-146
Tc00.1047053508799.280	Hypothetical protein, conserved.	6.60E-143
Tc00.1047053510103.30	Hypothetical protein, conserved .	2.90E-140
Tc00.1047053506725.70	Hypothetical protein, conserved|.	2.40E-124
Tc00.1047053507089.110	Hypothetical protein, conserved.	8.10E-124
Tc00.1047053508801.10	Hypothetical protein, conserved|.	9.20E-94
Tc00.1047053506947.10	Hypothetical protein, conserved|.	2.80E-93
Tc00.1047053509107.9	hypothetical protein, conserved|T.	5.30E-92
Tc00.1047053511639.20	Hypothetical protein, conserved|.	3.60E-91
Tc00.1047053510605.10	Hypothetical protein, conserved|.	6.60E-90
Tc00.1047053509669.160	Hypothetical protein, conserved.	9.70E-85
Tc00.1047053506175.100	Hypothetical protein, conserved.	1.20E-84
Tc00.1047053509537.90	Hypothetical protein, conserved|.	1.60E-80
Tc00.1047053504765.20	Hypothetical protein, conserved|.	2.80E-73
Tc00.1047053509431.10	Hypothetical protein, conserved .	8.00E-70
Tc00.1047053508799.290	Hypothetical protein, conserved.	2.70E-64
Tc00.1047053402857.10	Hypothetical protein, conserved|.	6.00E-52
Tc00.1047053510001.20	Hypothetical protein, conserved|.	1.40E-49
Tc00.1047053507529.20	Hypothetical protein, conserved|.	2.60E-48
Tc00.1047053504149.100	Hypothetical protein, conserved.	3.80E-36
Tc00.1047053506505.30	Hypothetical protein, conserved|.	1.90E-35
Tc00.1047053506175.110	Hypothetical protein, conserved.	2.40E-35
Tc00.1047053509669.170	Hypothetical protein, conserved.	1.00E-34
Tc00.1047053510001.31	Hypothetical protein, conserved|.	3.00E-34
Tc00.1047053507677.160	Hypothetical protein, conserved.	4.50E-34
Tc00.1047053507529.30	Hypothetical protein, conserved|.	6.60E-33
Tc00.1047053503837.10	Hypothetical protein, conserved|.	3.80E-32
Tc00.1047053507707.20	Hypothetical protein, conserved|.	3.80E-32
Tc00.1047053507383.10	Hypothetical protein, conserved|.	1.70E-09
Tc00.1047053508397.70	Transporter, putative|Trypanosom.	0.993
Tc00.1047053509875.240	Hypothetical protein, conserved.	0.9995

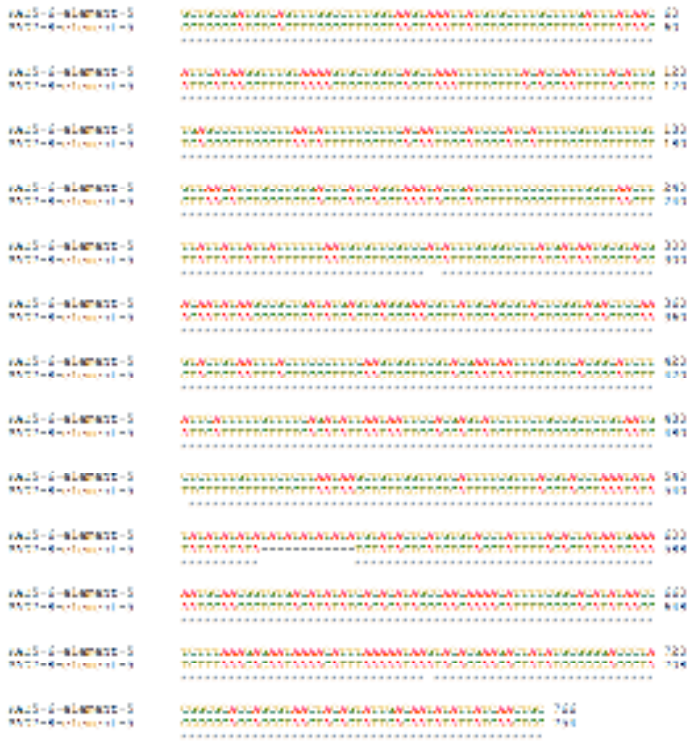


An alignment of the PAD intergenic elements. The PAD intergenic repetitive elements (A) 1, (B) 2 and (C) 4 were identified by multiple sequence alignments of the PAD intergenic regions.



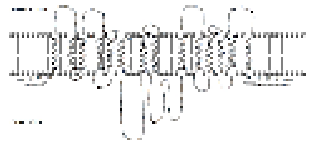
An alignment of the PAD intergenic elements The PAD intergenic repetitive element 3 was identified by multiple sequence alignments of the PAD intergenic regions. Note that element 3 is truncated downstream of PAD8 (PAD8-DOWN-element-3).

A



B

An alignment of the PAD intergenic elements The PAD intergenic repetitive elements (A) 5 and (B) 6 were identified by multiple sequence alignments of the PAD intergenic regions.



PAD1

PAD2

PAD3

PAD4

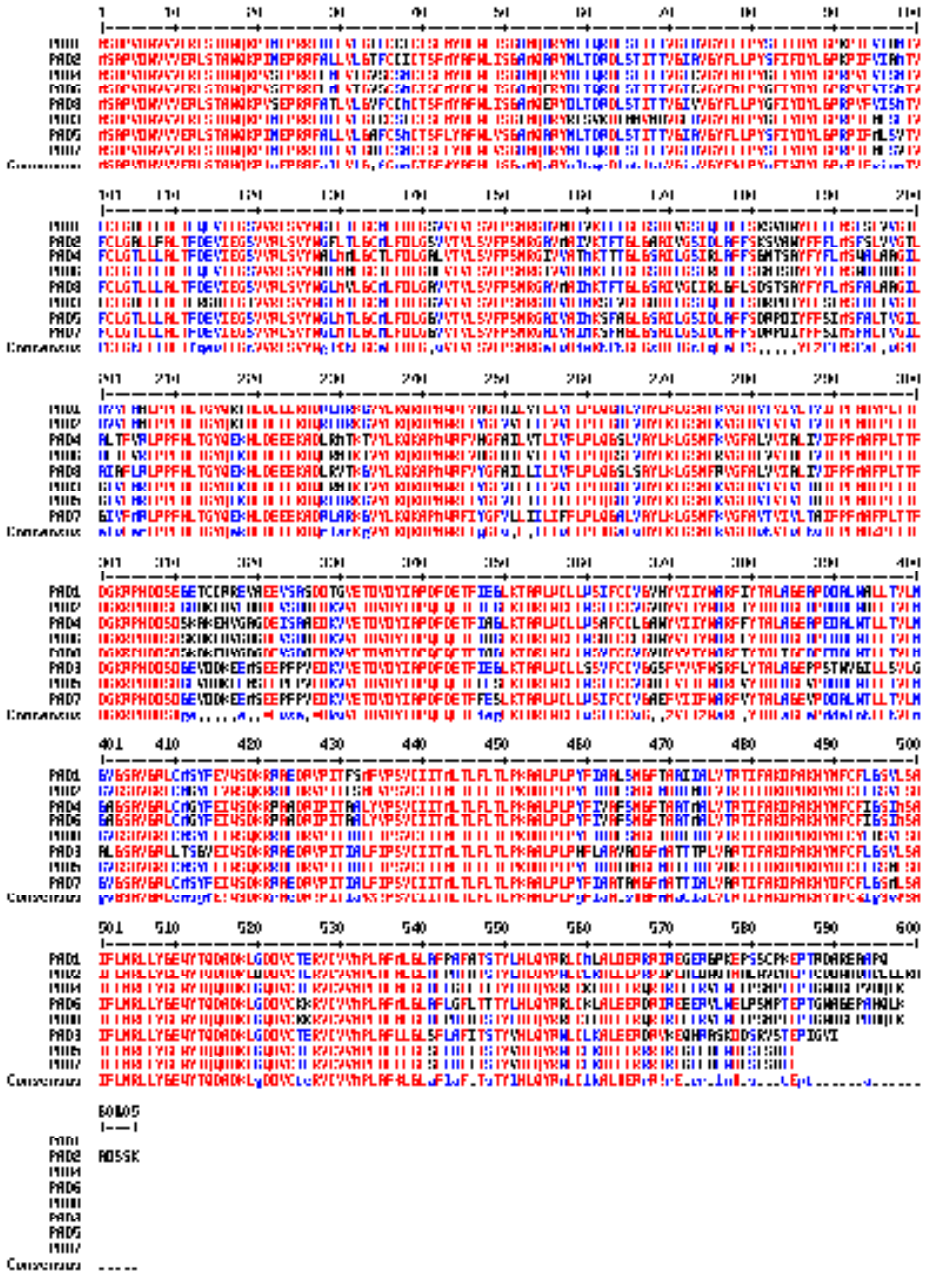
PAD5

PAD6

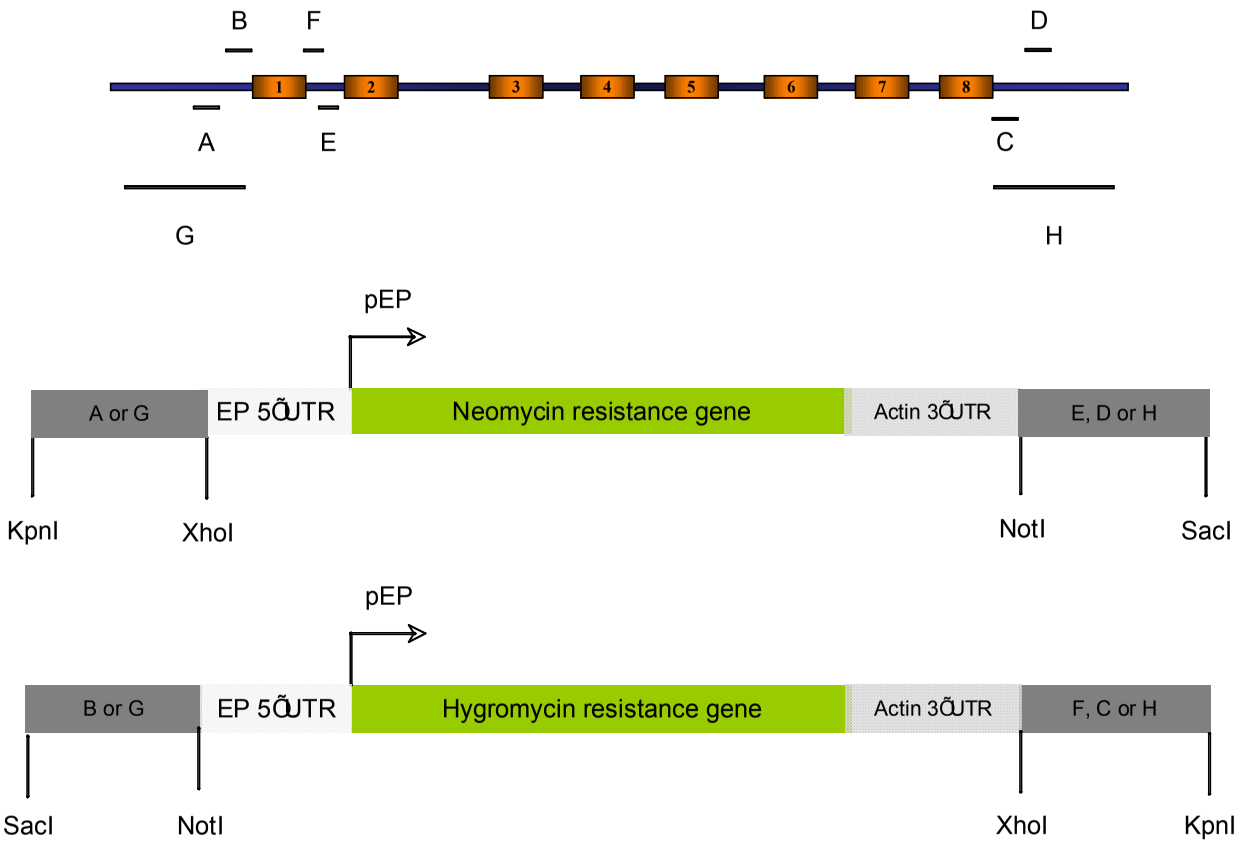
PAD7

PAD8

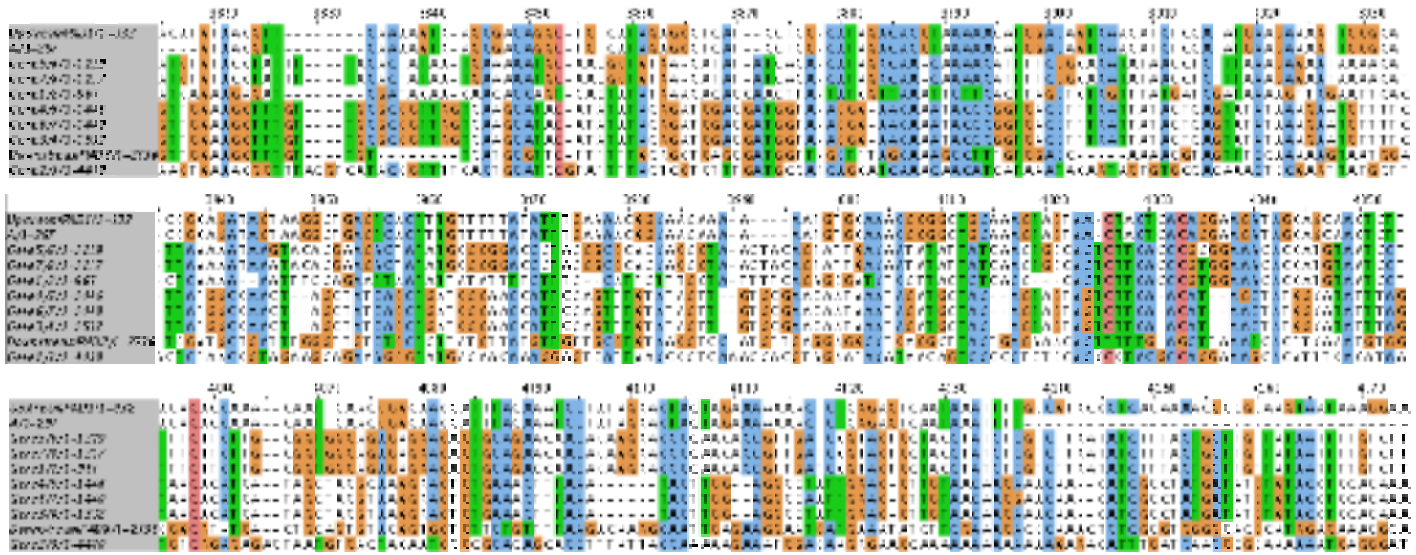
PAD protein secondary structures. The secondary structures of the PAD proteins were predicted as described in chapter 3. Note that all proteins share a similar predicted topology, with 14 trans-membrane spanning regions, and 2 large, intracellular loops.



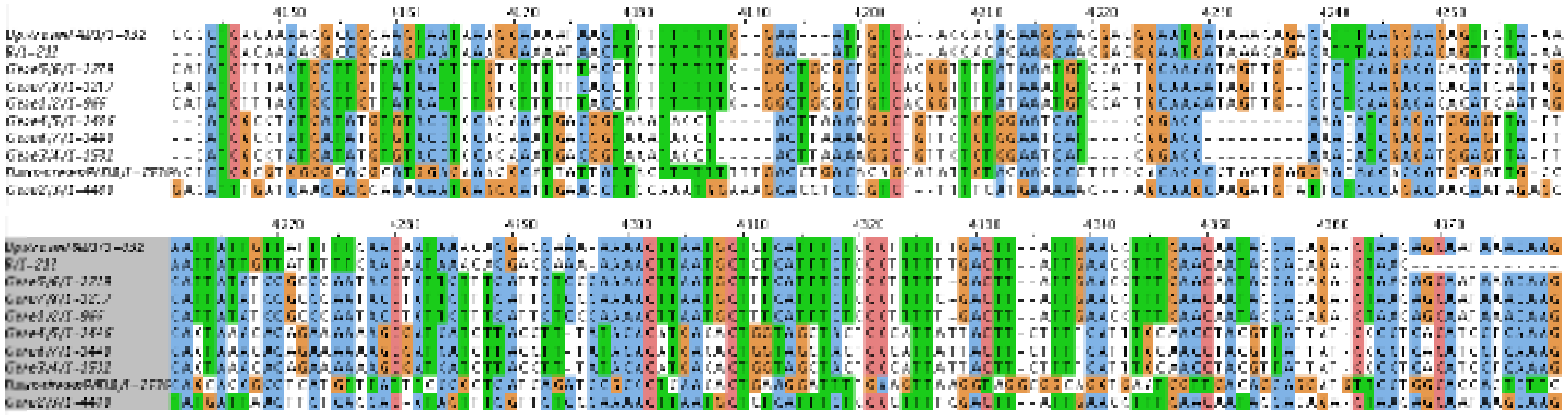
A PAD array protein alignment. The PAD proteins were aligned using Multalin to determine the conservation of amino acids between the proteins. Red indicates perfect conservation, blue indicates high conservation, and black indicates low conservation. Note that the C terminal end is the most heterogeneous



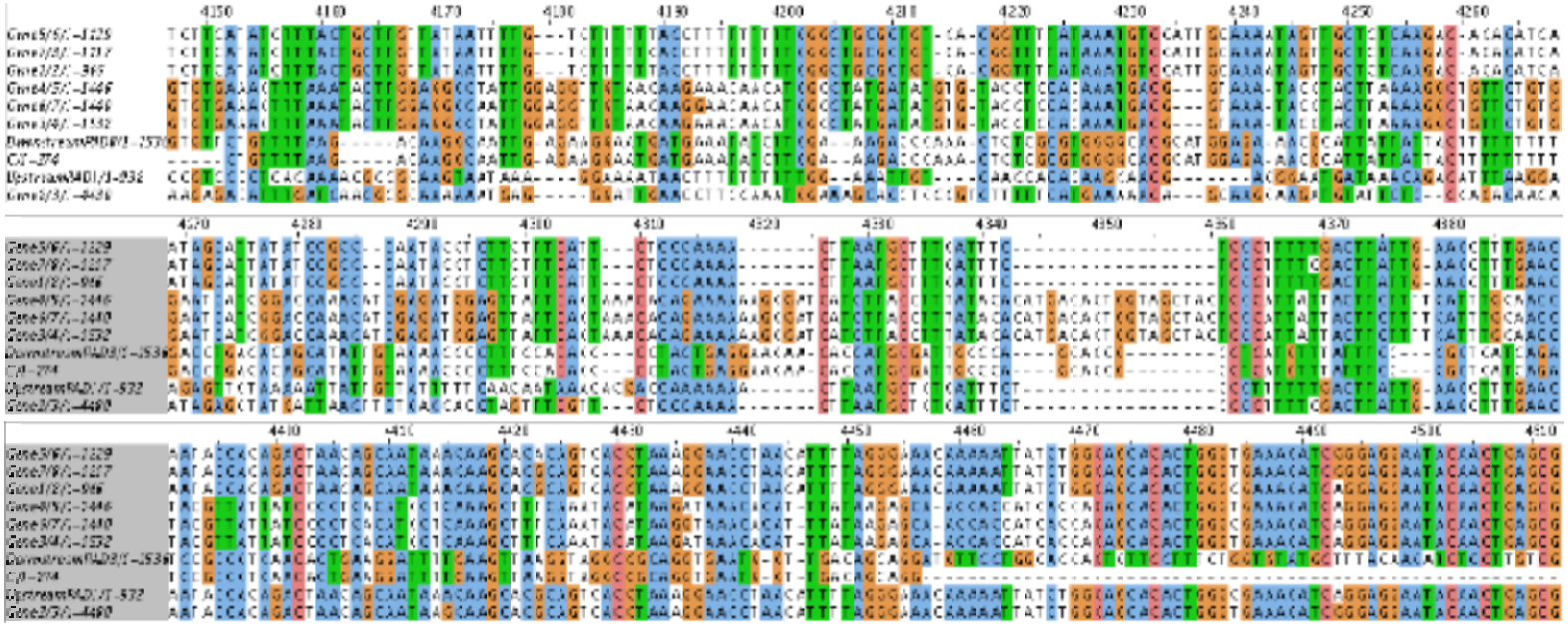
The target sequences are indicted on the PAD area of the genome (top). The neomycin (middle) and hygromycin (bottom) deletion constructs are shown as a chematic.



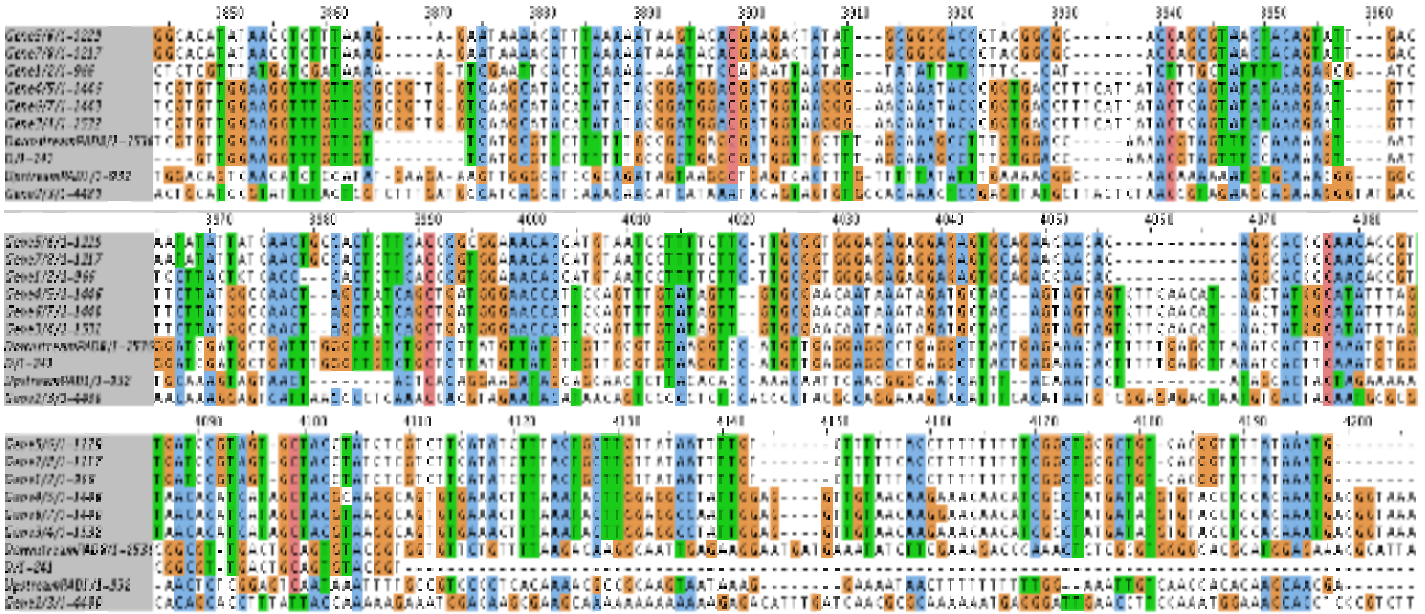
An alignment of target sequence A (A/1-297) with the PAD intergenic regions.



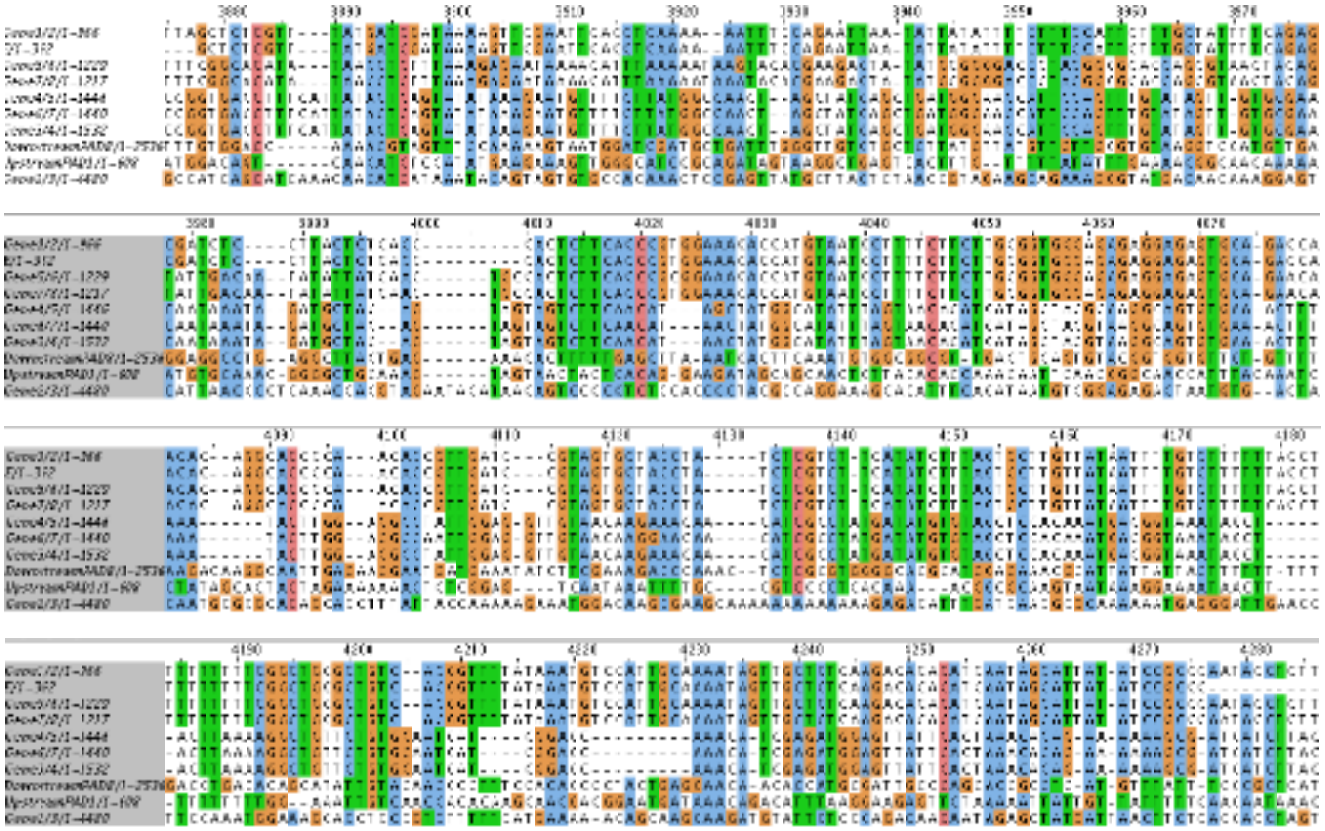
An alignment of target sequence B (B/1-212) with the PAD intergenic regions.



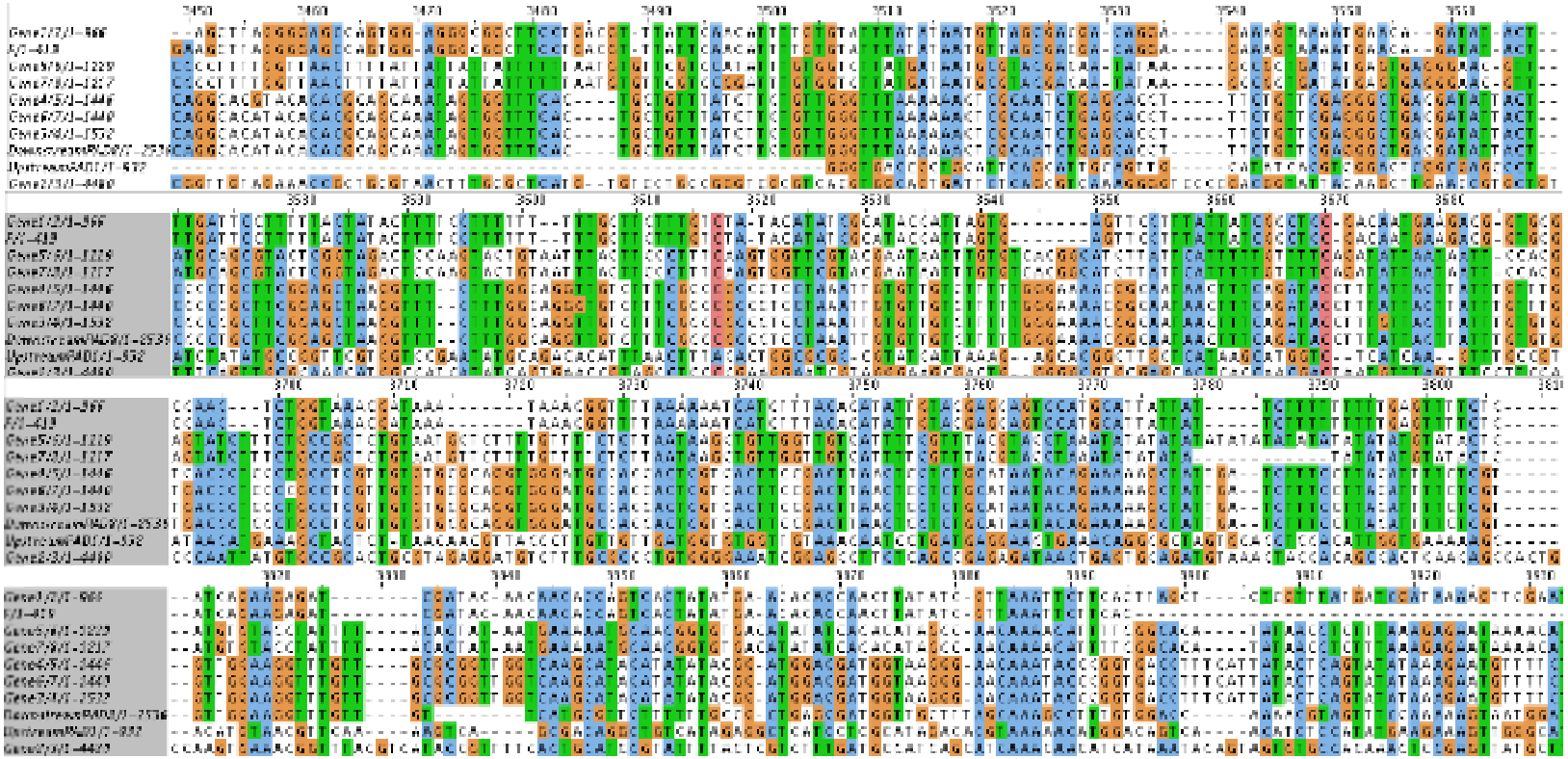
An alignment of target sequence C (C/1-274) with the PAD intergenic regions.



An alignment of target sequence D (D/1-241) with the PAD intergenic regions.



An alignment of target sequence E (E/1-362) with the intergenic regions.



An alignment of target sequence F (F/1-419) with the intergenic regions.